8-2011

GOING ALL THE WAY: PHYLOGEOGRAPHY AND TRANS-PACIFIC DIVERGENCE GENETICS OF NUCELLA LIMA

Lisa Cox
Clemson University, lncox@clemson.edu

GOING ALL THE WAY: PHYLOGEOGRAPHY AND TRANS-PACIFIC DIVERGENCE GENETICS OF *NUCELLA LIMA*

A Thesis
Presented to
The Graduate School of
Clemson University

In Partial Fulfillment
Of the Requirements for the Degree
Master of Science
Biological Sciences

By
Lisa Nicole Cox
August 2011

Accepted by:
Dr. Peter Marko, Committee Chair
Dr. Saara DeWalt
Dr. Margaret Ptacek
Dr. David Tonkyn

Abstract

Fluctuating climate over the last 2 million years (MY) has repeatedly caused latitudinal shifts in species distributions, fueling the hypothesis that the glacial-interglacial dynamics of the Pleistocene could have driven regional genetic differentiation and potentially speciation. For species whose distributions spanned the entire North Pacific, regional extinction of northern populations during cooler glacial periods may have resulted in isolation and genetic differentiation of eastern and western populations. To test this hypothesis, I gathered genetic data from a rocky shore intertidal gastropod, *Nucella lima*, whose current (i.e. warm interglacial) distribution spans the entire North Pacific. Mitochondrial DNA sequences are genetically structured with respect to eastern and western populations, suggesting an extended period of geographic isolation. Two additional nuclear genes also show greater differentiation across the Pacific than among populations on the same side. The population structure of these genes and the amount of genetic divergence and differentiation between the east and west indicate that *N. lima* persisted in refugia on each side of the Pacific Ocean during the Pleistocene. Indeed, multiple refugia may have existed in the western Pacific each with independent demographic histories subsequent to the last glacial maximum (LGM) due to limited genetic connectivity. Using the isolation-with-migration coalescent model, I have estimated divergence times between eastern and western populations that date within the early Pleistocene and shown that these sampled populations have experienced limited or absent gene flow since the LGM.

Dedication

I dedicate this manuscript to my family and friends who supported me throughout the research and writing process.

Acknowledgements

We would like to thank Ken Vicknair for help collecting samples and logistical support in Japan, the National Science Foundation for funding (NSF grants OCE-0550526 and OCE-0961996), Dr. Nadya Zaslavskaya for assistance traveling and collecting in Russia, and the members of the Marko Lab for their help and support and my thesis committee for reading and commenting on this document.

Table of Contents

Page

Title Page.....I

Abstract.....II

Dedication.....III

List of Figures III

List of Tables III

Background.....1

Introduction 6

Materials and Methods.....12

Results.....20

Discussion 46

Literature Cited 54

List of Figures

Page

Figure 1. Hypotheses....9
Figure 2. Sampling Map....13
Figure 3. Bayesian and Cladistic Trees....24
Figure 4. Parsimony Haplotype Networks....29
Figure 5. Isolation by Distance Graph....34
Figure 6. IMA Divergence Time Histograms....38
Figure 7. IMA Migration Rate Histograms....41

List of Tables

Page

Table 1. Summary Statistics....21
Table 2. COI Pairwise Phi-statistics....31
Table 3. SAMOVA Results....33
Table 4. Within Region IMA Nested Models....36
Table 5. IMA Results for eastern Pacific....37
Table 6. Trans-Pacific IMA Nested Models....39
Table 7. IMA Results for western Pacific....42
Table 8. IMA Results for trans-Pacific comparisons....44
Table 9. Beast Results....45

Background

Pleistocene Effects on Diversification

The Pleistocene occurred from 2.58 million to 11,800 years before present and was characterized by the latest period of frequent glacial-interglacial cycles. The climate cycles during this epoch elicited many responses from taxa affected directly and indirectly. Many taxi responded with range shifts (Bennet *et al.* 1991; Hewitt 2004; Lovette 2005) either in latitude or altitude. Altitudinal shifts were common in mountainous regions and may have been more localized than latitudinal shifts that could occur over large regions (Ehrich *et al.* 2007). Range shifts were evident in woody plant species that tracked suitable climate space as the climate changed while adapting to the local climate and different habitats (Davis & Shaw 2001). Latitudinal recolonization and expansion could have been rapid during the warmer interglacials and following the last glacial maximum (LGM) about 20,000 years ago. Recolonizations and expansions can be detected today by the loss of genetic diversity resulting from repeated bottlenecks and founder effects at the leading edge of colonization (Hewitt 1999).

Taxa survived through the glacial periods by persisting south of the ice sheets and permafrost, and in climatically suitable refugial areas within glaciated areas. Keppel *et al.* (2011) define a refugium as a suitable habitat that a taxon can retreat to, persist in, and potentially expand from when environmental conditions change and become suitable for habitation again. A single taxon may have persisted in several refugia that were isolated during glacial cycles and then expanded their ranges during interglacials, which re-established gene flow. In

some cases, populations that became fragmented into separate refugia may have become isolated, leading to separate genetic lineages as a result of vicariance (Carstens and Knowles 2007; Taberlet *et al.* 1998). The idea of vicariance in separate refugia first became prevalent in the late 1940s and led to the paradigm of late Pleistocene speciation. Rand (1948) introduced the idea that many avian species pairs may have diverged as a result of geographic isolation during periods of glaciation in the Pleistocene, and argued that this geographic pattern explained speciation despite the absence of current geographical barriers.

Responses to climate change can vary not only across taxa but also across regions. For example, *Arabis alpina,* an Arctic alpine and afro-alpine plant, is found across three different regions which primarily showed different responses to the glaciations in a wide variety of species: the North Atlantic region, which was recolonized after the last ice age; the European Alps, where range shifts were probably primarily altitudinal; and the high mountains of East Africa, where the contemporary mountain top populations result from range contraction (Ehrich *et al.* 2007). Ehrich documented that *A. alpina* showed different climate change responses in each region that corresponded to the expectations of each region given the environmental history and the evidence from other species. There is also evidence that speciation rates during the Pleistocene correspond to a spatial gradient where rates are higher in northern latitudes and areas directly affected by glaciers (Weir and Schluter 2004). Despite the fact that species respond similarly in a region, species have been shown to respond individualistically to

climate change rather than as distinct species assemblages (Jablonski and Sepkowski 1996; Marko *et al.* 2010; Roy et al. 1996). Woody plant species responded individualistically by range shifts to different regions, in different directions, and to different elevation ranges (Davis & Shaw 2001).

In recent years, the paradigm of late Pleistocene speciation has been challenged. For example, Klicka and Zink (1997) found that the speciation of 35 pairs of songbirds predate the Pleistocene. Speciation and extinction rates have failed to increase during times of environmental instability in some taxa lending credence to the paradox of the Pleistocene (Zink and Slowinski 1995; Coope 1995). However, most studies estimating speciation times only used sequence divergence from mitochondrial, or single-locus, data to date speciation, which is commonly known to overestimate sequence divergence. Multilocus datasets using model-based methods that take into account effective population size are necessary to accurately estimate time of divergence to understand the history of gene flow and population sizes.

In particular, the North Pacific Ocean has played a major role in the diversification of many taxa during the Pleistocene. Marine species distributed continuously or disjunctly across the North Pacific Ocean, called trans-Pacific taxa, were potentially isolated into eastern and western lineages during the glacial periods (Vermeij 1989). Eastern and western lineages of trans-Pacific taxa have been documented for species of different dispersal capabilities, particularly in different species of fish. For example, *Gadus macrocephalus,* Pacific cod, were isolated into separate glacial refugia, creating vicariance

between the northeastern and northwestern Pacific (Canino *et al.* 2010). Similarly, Grant and Utter (1984) found two distinct genetic lineages in Pacific herring, *Clupea pallasi*, corresponding to the eastern and western Pacific. Repeated glaciations in southern Alaska created a barrier to gene flow, leading to these distinct lineages. In the snailfish species complex, *Careproctus rastrinus*, multiple species were created in the Pleistocene, because of isolation in multiple marginal seas including the Sea of Japan, the Pacific coast of Japan, the Sea of Okhotsk, the Bering Sea, the Gulf of Alaska, and the Arctic Ocean (Kai *et al.* 2011). Harlin-Cognato (2006) also reported multiple lineages in the Stellar's sea lion that correspond to four glacial refugia south of the ice sheets that date within the Pleistocene. Although these studies demonstrate divergences that date within the Pleistocene, all divergences precede the LGM, which suggests that the more intense glaciations at the end of the epoch were not solely responsible for speciation but instead they probably reinforced the divergence between evolutionary lineages in isolated refugia that resulted from the earlier glaciations. In many cases, it is likely that secondary contact was not fully established between isolated populations during the interglacials or gene flow was limited between them if secondary contact occurred.

Diversification and speciation are common in the northwestern Pacific, perhaps more so than in the northeastern Pacific, because of the isolation of marginal seas that occurred when sea levels dropped. The narrowing and decrease in depth of the northern entrance isolated the Sea of Japan around 1.52 million years ago (mya), while temperature and salinity changes provided

different selection pressures (Kitamura *et al.* 2001). Kokita and Nohara (2010) studied *Leucopsarion petersii*, an ice goby that diverged into two major lineages from the Japan Sea and the Pacific Ocean in the late-early to middle Pleistocene. *Mugil cephalus*, a cosmopolitan flathead mullet, is composed of three cryptic species that are the result of isolation due to sea level and temperature fluctuations in the northwestern Pacific in the Plio-Pleistocene epochs with at least one species pair dating within the Pleistocene (Shen *et al.* 2011). The redlip mullet, *Chelon haematocheilus*, also show three spatially restricted lineages that result from isolation in marginal seas during low sea levels in the Pleistocene (Liu *et al.* 2007). In a non-fish example, an intertidal limpet, *Cellana nigrolineata,* forms three clades in Japan that date to Plio-Pleistocene with one split occurring 0.26-0.35 mya and the other occurring 1.28-4 mya (Nakano *et al*. 2010). Ilves and Taylor (2007) noted the importance of the northwestern Pacific in generating diversity before the Pleistocene climate fluctuations in the smelt genus *Hypomesus.* The most recent split occurred in the Pliocene to early Pleistocene, which suggests a historical role of this region in generating species diversity. Speciation in the northeastern Pacific is not as well documented, but did occur, especially in species with limited dispersal capability, such as the seastar genus *Leptasterias*, which contains four species pairs with divergence times from 0.5-1.2 mya (Foltz *et al.* 2008).

In conclusion, it may be premature to state whether the paradigm of late Pleistocene speciation has sufficiently been challenged. The generation of diversity has both coincided with the timing of the glacial cycles as well as before

the Pleistocene. Studies that only used sequence divergence analyses from a single locus to place a date on species divergence may not be reliable and therefore should be treated with caution until more sophisticated analyses such as model-based methods using multilocus datasets can confirm these earlier estimates. More extensive research is needed to adequately evaluate the relevance of this paradigm.

Introduction

A long-standing debate in the fields of biogeography and evolutionary biology is if the demographic impacts of late Pleistocene glacial cycles provided opportunities for geographic isolation and an increase in species diversity (Valentine & Jablonski 1993). Frequent glacial-interglacial cycles over the last 2 million years (MY) caused latitudinal shifts (Bennet *et al.* 1991, Lovette 2005) and potential fragmentation of species' distributions. Although many Northern Hemisphere species' assemblages have been shown to have shifted their ranges southward during glacial periods and subsequently, recolonized higher latitudes in a similar fashion (Addicot 1966; Hewitt 2004), there is evidence that species respond individualistically to climate change rather than as distinct assemblages of co-distributed species (Jablonski and Sepkowski 1996; Roy et al. 1996; Davis & Shaw 2001; Marko *et al.* 2010).

The paradigm of late Pleistocene speciation suggests that the more intense glacial cycles of the late Pleistocene should have generated more allopatric speciation than earlier in the Pleistocene. Thus, many present-day species pairs would have dates of divergence within the late Pleistocene (Rand

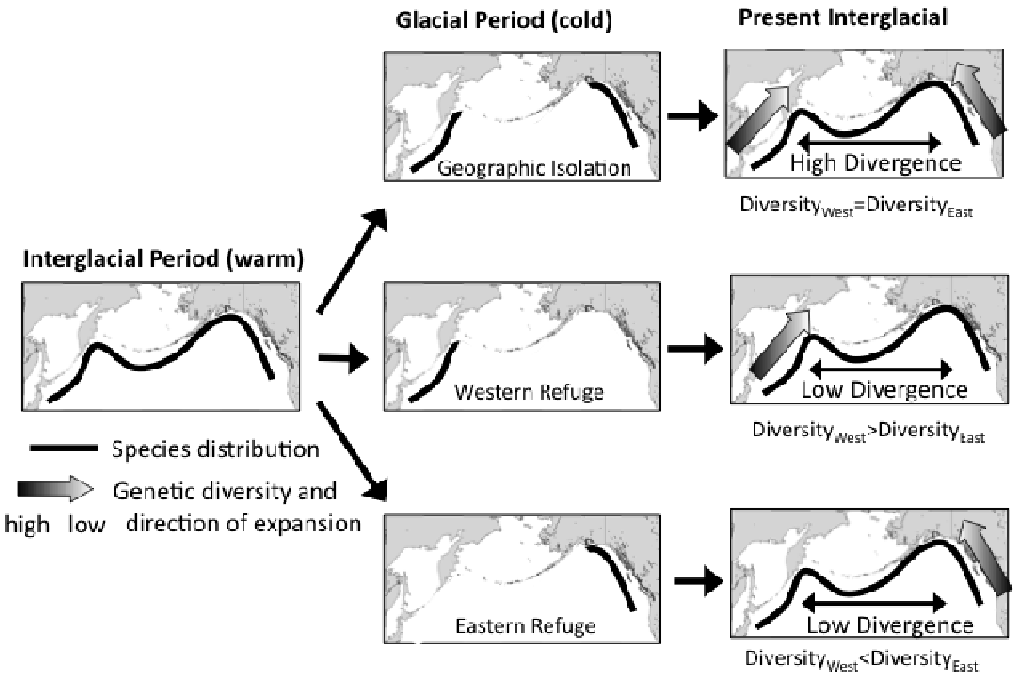
1948). This paradigm is challenged, however, by evidence in North American avian species that differentiation is too great between sister species to have occurred as recently as the late Pleistocene. Instead, the origin of much of the current avian diversity dates to within the Pliocene (Klicka & Zink 1997). There are a multitude of studies supporting divergence that pre-dates the Pleistocene in many species pairs (e.g., Cracraft & Prum 1988; Bush 1994; Riddle 1996). However, because only single locus sequence divergences are used to estimate divergence times and do not take into account current or ancestral population processes, it is likely that these speciation time estimates could be inaccurate and potentially overestimated (Arbogast *et al.* 2002). In contrast, Weir and Schluter (2004) show evidence of a latitudinal effect in speciation rates during the Pleistocene. Speciation rates were higher in boreal habitats directly fragmented by glaciers than in lower latitude regions, documented particularly in avian species. Because biogeographic regions or habitats may play a role in the speciation rates seen as a result of climate change, it is important to evaluate this paradigm across a multitude of habitats and in a diversity of taxa.

Although much of the debate surrounding the role of climate change has focused on boreal habitats and birds, other habitats were affected heavily by the fluctuating climate of the Pleistocene as well, such as marine intertidal habitats. Not only were some intertidal habitats lost or changed structurally by the advancing and retreating of the terrestrial ice sheets, but sea levels, salinity and temperatures fluctuated as well (Clark *et al.* 1999; Mitrovica 2003; Rohling *et al.* 1998; Tanaka & Takahashi 2005). In the North Pacific, during cooler Pleistocene

glacial periods, regional extinction of northern populations caused by widespread glaciation would have divided trans-Pacific nearshore marine taxa into isolated eastern and western populations (Vermeij 1989). About 10% of molluscan species in the northwestern Pacific are estimated to have been derived from vicariant events such as these during the Pliocene, but it is less well known whether these speciation events continued into the Pleistocene (Coan 1971; Lindberg 1982; Matsukuma 1986; Lindberg and Marincovich 1988; Vermeij 1989). Cryptic species are common in the North Pacific, and molecular data are important in establishing the presence of cryptic species as well as distinguishing them (Knowlton 1993; Marko 1998). For these North Pacific taxa, what little molecular data exist suggest that east-west differentiation has occurred. For example, Canino *et al*. (2010) detected considerable divergence between northeastern and northwestern populations of *Gadus macrocephalus,* Pacific cod, using both microsatellite and mitochondrial DNA and they argued that at least two glacial refugia existed, one on each side of the North Pacific. Another example of east-west genetic divergence resulting from restricted gene flow due to repeated glaciation in southeast Alaska can be found in Pacific herring, *Clupea pallasi* (Grant & Utter 1984).

Different biogeographic responses to Pleistocene climate change should leave behind different characteristic patterns of genetic diversity (Figure 1). For example, if a taxon was isolated into eastern and western populations during glacial periods and each population experienced different evolutionary histories, genetic divergence should be higher between rather than within regions and

Figure 1. Genetic patterns expected under different Pleistocene scenarios for a trans-Pacific taxon.



genetic endemism may be present. This scenario may also result in a latitudinal gradient of genetic diversity (Vucetich & Waite 2003) in which diversity is highest at the southern ends of the ranges on each side of the Pacific, where the putative glacial refugia were located, and decreases northwards in the direction of expansion and recolonization. This expectation of leading edge colonization has been documented in other taxa that expanded and shifted their ranges after the last glacial maximum (LGM) approximately 20,000 years ago (Schmitt 2007). Alternatively, cooler Pleistocene glacials may have resulted in geographic restriction to a single refuge on either side of the Pacific (Vermeij 1989). Populations living now in either an eastern or western refuge should have higher genetic diversity relative to surrounding areas and show greatly decreased genetic diversity in the direction of expansion and recolonization.

To investigate the biogeographic and evolutionary impacts of Pleistocene climate change on patterns of species diversity, I conducted a phylogeographic study of the rocky shore gastropod *Nucella lima*, which currently has a continuous distribution across the entire North Pacific from Hokkaido, Japan to Northern Vancouver Island, Canada (Vermeij et al. 1990; Figure 2). This predator is found in the mid-low intertidal and has benthic encapsulated development in which juvenile snails emerge directly from egg capsules. In a previous study, four haplotypes sampled from *N. lima* showed 1.8% mitochondrial DNA (mtDNA) sequence divergence between populations in Russia and Alaska (Collins *et al.* 1996), corresponding to a split no earlier than the Pleistocene (assuming a 1-2%/MY rate of divergence) but also exceeding the divergence between other

cryptic but reproductively isolated species of *Nucella* (Palmer *et al.* 1990; Marko 1998). This pattern of sequence divergence suggests evidence of northern extinction of populations during the glacial periods with persistence on each side of the Pacific Ocean, but because these sample sizes are not sufficient to capture the full extent of diversity, more extensive data from this species may provide an example of speciation coincident with the climate cycles of the Pleistocene. Samples from more localities and loci are also needed to understand the history of gene flow and isolation within and between eastern and western populations. Isolation with migration models also need to be tested to distinguish between shared genetic variation as a result of ancestral polymorphism or current gene flow, while also taking into account other population parameters such as mutation rate and population sizes and, therefore, more accurately conclude whether the timing of this split occurred during the Pleistocene (Marko and Hart 2011).

In order to further investigate whether the Pleistocene climate cycles generated diversity in the northern Pacific Ocean, I used molecular genetic methods to reconstruct the biogeographic responses of a trans-Pacific rocky-shore predatory gastropod, *N. lima*, to Pleistocene glacial cycles. With extensive sampling across their entire geographic distribution, I collected data from one mitochondrial and two nuclear genes in order to examine molecular diversity and differentiation patterns and estimate times of divergence using coalescent and model-based population genetic analyses.

Materials and Methods

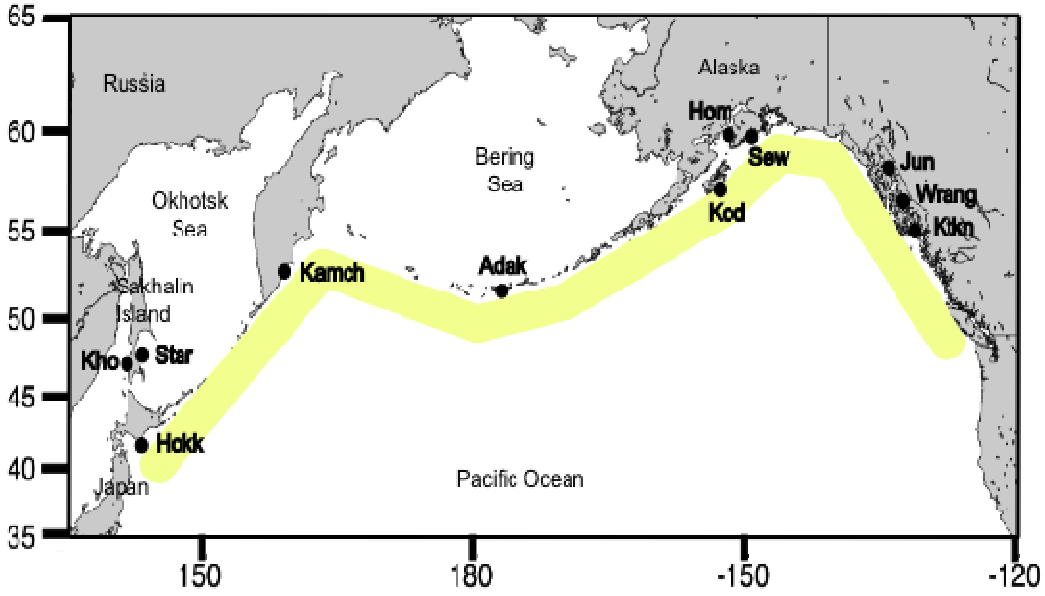
Sampling

Nucella lima was collected from 11 sites spanning the entire known distribution of this species across the North Pacific Ocean (Figure 2) between 2003 and 2008. The number of snails collected at each site ranged from 2-23 individuals (Table 1). Eastern samples included Kodiak, Homer, Seward, Juneau, Wrangell, and Ketchikan and western samples included Kamchatka, Kholmsk, Starodubskoye and Hokkaido. Because only two individuals were collected from Adak, these samples were excluded from most population genetic analyses and were not classified as either eastern or western Pacific. Genomic DNA was extracted from foot tissue using a standard CTAB protocol (Grosberg et al. 1996), resuspended in water, and stored at -20℃.

Mitochondrial DNA

I amplified cytochrome c oxidase subunit 1 (COI) using the polymerase chain reaction (PCR) with the primers, LCOI-nlF1 (5'-TAGGGCAACCAGGAGCTTTA-3') and HCOI-nlR1 (5'-AAGATAGGATCACCACCTCCTG-3'). I directly sequenced 504 base pairs (bp) of COI from a total of 119 individuals from all 11 sites. Contigs were assembled and edited in SEQUENCHER 4.8 (Gene Codes Corp., Ann Arbor, MI, USA) and manually aligned in SE-AL v2.0a11 (Rambaut 2002). Repetitive sequences were collapsed into representative haplotypes using COLLAPSE v1.2 (Posada 2006). Unique haplotypes were deposited in GenBank (Accession numbers: JN049611-JN049630).

Figure 2. Map of sampling localities for *Nucella lima*. Map was created with OMC (http://www.aquarius.ifm-geomar.de/make_map.html).



Nuclear DNA

Two nuclear loci were sequenced in order to provide multilocus estimates of population parameters in coalescent-based analyses. I targeted six populations for this additional sampling: Hokkaido, Starodubskoye, and Kamchatka from the western Pacific and Kodiak, Wrangell, and Ketchikan from the eastern Pacific (Figure 2). I amplified 614 bp of the ITS region, which comprises the internal transcribed spacers (ITS1 and ITS2) and the 5.8S rRNA gene, using the forward and reverse primers from Insua *et al.* (2003). I also amplified 817 bp of elongation factor 1-α (EF1a) with primers EF-F8 (5'-CAG ACT CGT GAA CAC GCT-3') and EF-R792 (5'-AGA CAT CCT GAA GGG GCA-3'). PCR products for both nuclear genes were cloned with the pGEM®-T Easy Vector System using Electromax DH10B *E. coli* cells (Promega). I sequenced 614 bp of the ITS region for 38 individuals and 817 bp of EF1a were sequenced for 33 individuals from the aforementioned six sites. Sequences from both loci were edited in SEQUENCHER and initially aligned in CLUSTALX v1.83.1 (Chenna *et al.* 2003) and then checked by eye. Nuclear sequences were also collapsed into representative haplotypes in COLLAPSE. The full sequences of unique haplotypes including recombinant sites were deposited in GenBank (Accession numbers: XXXXXXXX-XXXXXXXX).

Genetic diversity

Because population genetic analyses assume no recombination within a locus, I removed probable recombinant sites from ITS and EF1a using IMGC (Woerner *et al.* 2007), which retains the largest non-recombining block of DNA sequence

using violations of the 'four gamete test' (Hudson and Kaplan 1985). I performed most analyses using the full dataset as well as this non-recombining block of sequences and some analyses using only the non-recombining block of sequences. MODELTEST v3.8 (Posada 2006; Posada & Crandall 1998) determined the best-fit DNA substitution model for each of the three loci under the Bayesian Information Criterion (BIC). The best-fit model for COI in MODELTEST was HKY+I, and the best fit model for both EF1a and ITS was HKY.

Standard molecular diversity statistics were calculated in ARLEQUIN v3.11 (Excoffier *et al.* 2005) including nucleotide and haplotype diversity, Tajima's (1989) *D* and Fu's (1997) *FS* for both the full dataset and with possible recombinant sites removed. Significance levels of Tajima's *D* and Fu's *FS* were tested in ARLEQUIN using 10000 permutations. Fu's *F*S tends to be more sensitive than Tajima's *D* (Ramos-Onsins & Rozas 2002), the power of which tends to be limited to particularly strong or recent selection, or demographic changes (Hedrick 2005).

Gene trees

I constructed gene genealogies for the full dataset of each locus using three methods. First, a bootstrapped (1000 replicates) maximum parsimony (cladistic) analysis was performed for each locus with PAUP* v4.0 (Swofford 2001). Next, I constructed statistical parsimony networks for each locus in TCS v1.18 (Clement *et al.* 2000). COI and EF1a were set with a 95% connection limit, while ITS was set with a 90% connection limit in order for all haplotypes to fall

into a single parsimony network, and gaps were coded as missing data. Lastly, I examined phylogenetic relationships using Bayesian inference with MRBAYES v3.1 (Huelsenbeck & Ronquist 2001) applying the GTR+G+I model. The Bayesian searches included a uniform prior distribution of parameters and four simultaneous chains run for 5×10^6 generations, sampling trees every 100 generations. A burn-in period was set to discard 2500 trees before computing a consensus tree of all compatible groups with nodal support given by the posterior probability of each clade. I assessed convergence by confirming that the average standard deviation of split frequencies was below 0.01 after all generations had completed and repeating the analysis three times. For the Bayesian and cladistic analyses, only a single representative of each haplotype was included. *Nucella lamellosa* and *Acanthinucella spirata* were used as outgroups in the Bayesian and cladistic analysis of COI; *A. spirata* was used for ITS and the sister taxon of *N. lima*; and *N. freycinetti* was used for EF1a.

Population structure

An analysis of molecular variance (AMOVA) was calculated in ARLEQUIN at each locus for all localities sampled using 10000 permutations. I also performed three Mantel tests in ARLEQUIN for each locus to assess isolation by distance (IBD) (Mantel 1967) and migration-drift equilibrium: one test including all samples, one test including only western Pacific samples, and one test including only eastern Pacific samples. I calculated geographic distance in Google Earth (http://earth.google.com/) as the shortest oceanic distance between two sites.

To test for geographic consistency in genetic structure across all three loci, I used SAMOVA v1.0 (Dupanloup *et al*. 2002) to define geographically homogenous groups that are maximally differentiated from each other using a simulated annealing approach. The Tamura DNA substitution model was used as the most appropriate model available from ModelTest for all three loci because, like the HKY model, it allows unequal nucleotide frequencies and distinguishes between the rate of transitions and transversions. Because I only had multilocus data for six populations, I tested k=2 through k=5 for each locus, where k equals the number of groups. The value of k that maximizes among group differentiation (Φ_{CT}) and reduces within group differentiation (Φ_{SC}) is the number of panmictic groups that best fit the data.

Population divergence models

I used IM$_A$ (Hey & Nielson 2004) to fit a coalescent model of isolation with migration to data for each pair of sample sites. This approach distinguishes shared polymorphisms due to common ancestry from those due to migration and thus prevents underestimation of divergence time. The only samples used for this analysis were those for which I had data for all three loci, which included three localities from each side of the northern Pacific Ocean for a total of 15 pairwise population comparisons. This model allows estimation of multiple parameters, including genetic diversities, θ_1, $\theta_{2,}$ and θ_A, for each extant population and the ancestral population; migration rate into each population, m_1 and $m_{2;}$ and the time of divergence, *t*. Because this analysis assumes no recombination, only the

dataset excluding recombinant sites was used. The inheritance scalar was set to 0.25 for the mitochondrial locus to reflect the reduced effective population size due to the maternal, haploid inheritance of mtDNA and was set to 1.0 for the nuclear loci.

Preliminary runs of IMA in "MCMC Mode" or M-mode were used to optimize and set upper prior probabilities for the parameters. IMA was then run in "Load Trees Mode" or L-mode to determine if a simpler model had a higher likelihood than the full-parameter model using likelihood ratio tests. Specifically, I was interested if a model with no gene flow could be rejected or was equally probable to the full model with gene flow. If L-mode was unable to reject a model with no migration, then migration was set to zero for subsequent M-mode runs in order to narrow estimates of the other parameters. Three independent runs started with a different random number seed in M-mode were conducted for each analysis to assess convergence in IMA. All runs included at least 10 million steps, and effective sample sizes and parameter trend plots were monitored to make sure the parameter space was being fully explored. Because IMA only requires a mutation rate for one locus in multilocus datasets, a genus specific mutation rate (μ) of 3.83×10^{-6} substitutions/locus/year was specified for COI only (McGovern *et al.* 2010). Because this clock reflects a substitution rate, divergence times and population sizes could be overestimated. IMA uses this single mutation rate to calculate a mean mutation rate across all loci and then uses this mean rate to convert parameter estimates to biologically informative estimates. If posterior probabilities did not produce a clear peak in likelihood or did not drop to zero and

confidence intervals were unclear, results were unreliable and, therefore, not provided.

Population size change

I also used BEAST v1.6.1 (Drummond & Rambaut 2007) to characterize population size change over time by generating an Extended Bayesian Skyline Plot (EBSP) (Drummond *et al*. 2005) for the non-recombining dataset. I generated an EBSP for each collection locality using an MCMC run of 500 million steps. Each locality that was analyzed showed evidence from other analyses for restricted gene flow to other populations to satisfy the assumption that the haplotypes were sampled from a single population. For three populations, one locus had to be omitted from each because it was not variable in that population and thus, not informative for BEAST. Kamchatka and Kodiak were run including only ITS and EF1a, and Ketchikan was run including EF1a and COI. Each analysis was repeated with suggested alterations of priors until all effective sample sizes (ESS) were greater than 100. For each population, the marginal posterior distribution of the growth rate was evaluated to determine if constant population size could be rejected based on the exclusion of zero in TRACER v1.5 (Rambaut & Drummond 2009). I also compared the EBSP model to two simpler models: constant population size and exponential growth. The two simpler demographic models were set up in the same way as the EBSP runs. Comparisons were carried out using Bayes Factor Tests in TRACER using importance sampling of the marginal likelihoods of each of the models calculated

using the method from Newton and Raftery (1994) with the modifications proposed by Suchard *et al.* (2001). Guidelines from Jeffreys (1991) were used to evaluate strength of support for each model.

Results

Genetic diversity

Of the 119 samples sequenced for COI, there were 20 haplotypes, 11 of which were found in Hokkaido, Japan (Hokk), the most diverse location I sampled. Other sites contained only one to three haplotypes. Across all localities, nucleotide diversity ranged from 0 to 0.0179. Although the range of nucleotide diversities within the east and west overlapped considerably, the average nucleotide diversity across population means was about six times higher in the west (0.0026) than in the east (0.000433) (Table 1). The haplotype diversity across all sites ranged from 0 to 1. Similarly, the average haplotype diversity in the west (0.4175) was slightly higher than in the east (0.2117) (Table1). Adak had the largest nucleotide and haplotype diversities, but this sample consisted of only two sequences, one from each of two highly divergent eastern and western haplotype clades (see below).

For the nuclear loci, nucleotide diversities were similar for EF1a and ITS (Table 1). Haplotype diversities were higher for EF1a than for ITS, although the ranges in estimates overlapped. For ITS, just as in COI, the average nucleotide and haplotype diversities were slightly larger in the west (π_{AVG}= 0.0175 and h_{AVG}= 0.89) than in the east (π_{AVG}= 0.0026 and h_{AVG}=0.80) although more distinctly in

Table 1. Molecular diversity statistics for all collecting sites: sample size (*n*), number of unique haplotypes (n_h), nucleotide diversity (π), gene diversity (*h*), Tajima's D, Fu's F_S

Locus		Location	*n*	n_h	π	*h*	D	Fs
COI	West	Hokk	23	11	0.0044	0.89	-0.37	-4.91*
		Kho	5	2	0.0040	0.40	-1.12	2.64
		Star	10	3	0.0020	0.38	-1.74*	0.48
		Kamch	4	1	0.0000	0.00	n/a	n/a
		Adak	2	2	0.0179	1.00	0	2.19
		Kod	11	1	0.0000	0.00	n/a	n/a
		Hom	8	3	0.0010	0.46	-1.31	-0.99*
		Sew	15	1	0.0000	0.00	n/a	n/a
		Jun	12	1	0.0000	0.00	n/a	n/a
		Wrang	15	2	0.0008	0.42	0.74	0.91
	East	Ktkn	14	3	0.0008	0.39	-0.96	-0.85
ITS	West	Hokk	7	7	0.0008	1.00	-0.88	-3.08*
		Star	6	5	0.0459	0.93	0.88	2.57
		Kamch	6	3	0.0059	0.73	0.85	2.16
		Kod	6	5	0.0034	0.93	0.31	-2.04*
		Wrang	7	6	0.0032	0.95	-0.27	-3.41*
	East	Ktkn	7	3	0.0013	0.52	0.55	-0.44
EF1a	West	Hokk	5	5	0.0030	1.00	-0.17	-2.68*
		Star	6	6	0.0076	1.00	-1.27	-1.79
		Kamch	6	6	0.0173	1.00	0.48	-0.47
		Kod	7	6	0.0031	0.95	-0.04	-2.71*
		Wrang	6	5	0.0023	0.93	-1.23	-2.26*
	East	Ktkn	7	4	0.0038	0.71	-1.13	0.69

* indicates *P*-value < 0.05

nucleotide diversities. EF1a showed the same trend; average nucleotide diversities were 0.0093 and 0.0031 and average haplotype diversities were 1.0 and 0.863 in the west and east, respectively. The data with potential recombinant sites removed showed similar patterns of higher diversity in the west, but overall diversity statistics were lower (data not shown).

Tajima's *D* was negative at more than half of the estimates, but only one (Starodubskoye at COI) was statistically different from zero which indicates a population expansion or purifying selection (Table 1). Fu's F_S were also negative at more than half of the sample sites, but more estimates were found to be significantly negative across loci (Table 1). The nuclear loci were consistent in that Hokkaido, Kodiak, and Wrangell had significantly negative Fu's F_S, indicating Northeastern sites were more likely to deviate from neutrality than sites in the west Pacific. Overall, Tajima's *D* and Fu's F_S were not consistent with each other but were not necessarily contradictory because most estimates did not differ significantly from zero.

Gene trees

The cladistic and Bayesian COI gene trees were similar in topology and geographic distribution of alleles. In addition, clades were supported with high posterior probabilities and bootstrap support for both trees (Figure 3a). For example, all Alaskan samples composed a single, monophyletic clade with high support estimates (posterior probability = 0.97, bootstrap percentage = 95). Starodubskoye (Star) and Kholmsk (Kho), which were located on Sakhalin

Island, also formed a well-supported monophyletic clade that was the sister-group to all other *N. lima* samples (posterior probability = 0.63, bootstrap percentage = 99). All of the individuals from the other Russian sample site, Kamchatka, shared a single haplotype and were nested within the Hokkaido samples. Adak shared one haplotype with the samples from Kamchatka and one haplotype with the most common Alaskan haplotype, which was found at all Alaskan sampled localities. The Hokkaido samples were diverse and paraphyletic with both Kamchatka and Alaskan samples nested within the clade.

The ITS region topology did not show strong geographical structuring compared to COI (Figure 3b) with respect to eastern and western samples. *Nucella lima* was not supported as a monophyletic group with respect to the outgroup due to two divergent haplotypes from Starodubskoye on Sakhalin Island that formed a well-supported clade (posterior probability = 0.97, bootstrap percentage = 99). Of the four clades present in the tree, only two were restricted geographically, one being the two Sakhalin haplotypes and the other was an Alaskan clade. Support for this clade was fairly high (posterior probability = 0.73, bootstrap percentage = 62), but the Alaskan samples were not monophyletic because haplotypes outside of this clade were found at the Alaskan sites. The other two clades contained haplotypes from several populations spanning large geographic distances. Also, in a couple of cases, haplotypes were shared across these same geographic distances. For example, haplotypes ITS_M1 and ITS_M2 were shared across Japan, Russia, and Alaska.

Figure 3. Bayesian consensus tree for *N. lima* generated in MrBayes v3.1.2 (Huelsenbeck and Ronquist 2001) using a) COI, b) ITS, and c) EF1a markers. Posterior probabilities are displayed on the nodes, followed by the bootstrap support estimates. Labels on the terminal branches include haplotype name followed by the populations that haplotype was found in coded in the same format as Figure 2 and the frequency if greater than one. Codes in the haplotype type have the following meanings: RU-Russia; JP-Japan; AK-Alaska; M-mixed regions.

Figure 3a.

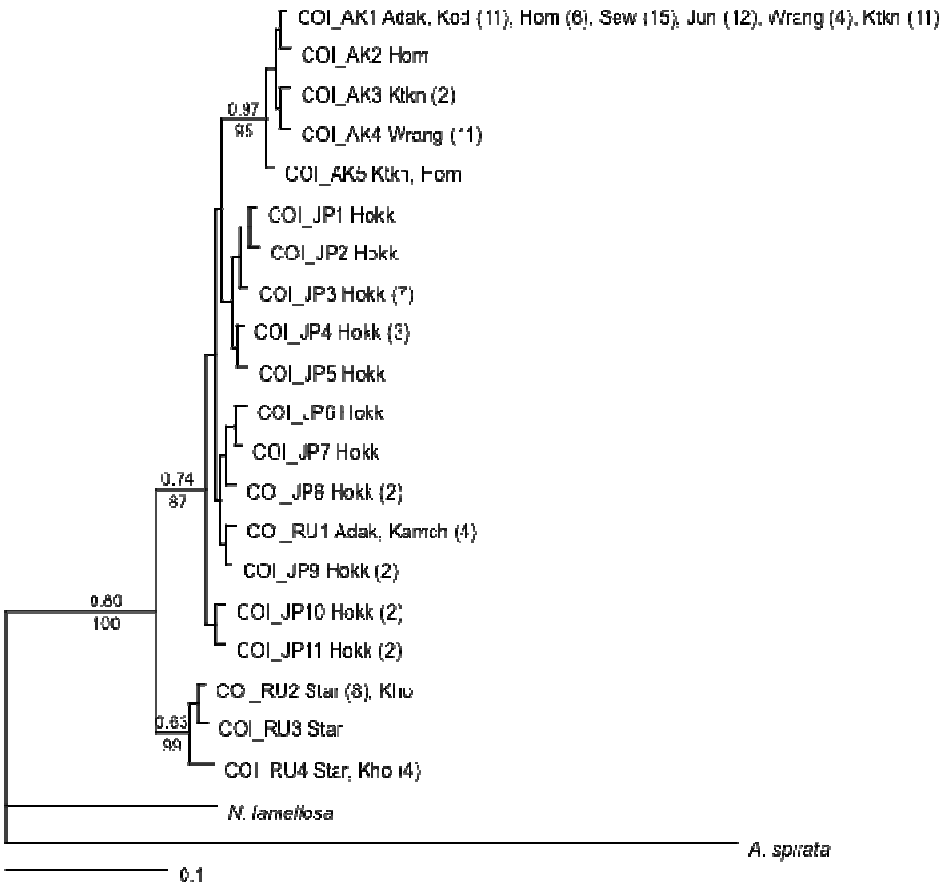


Figure 3b

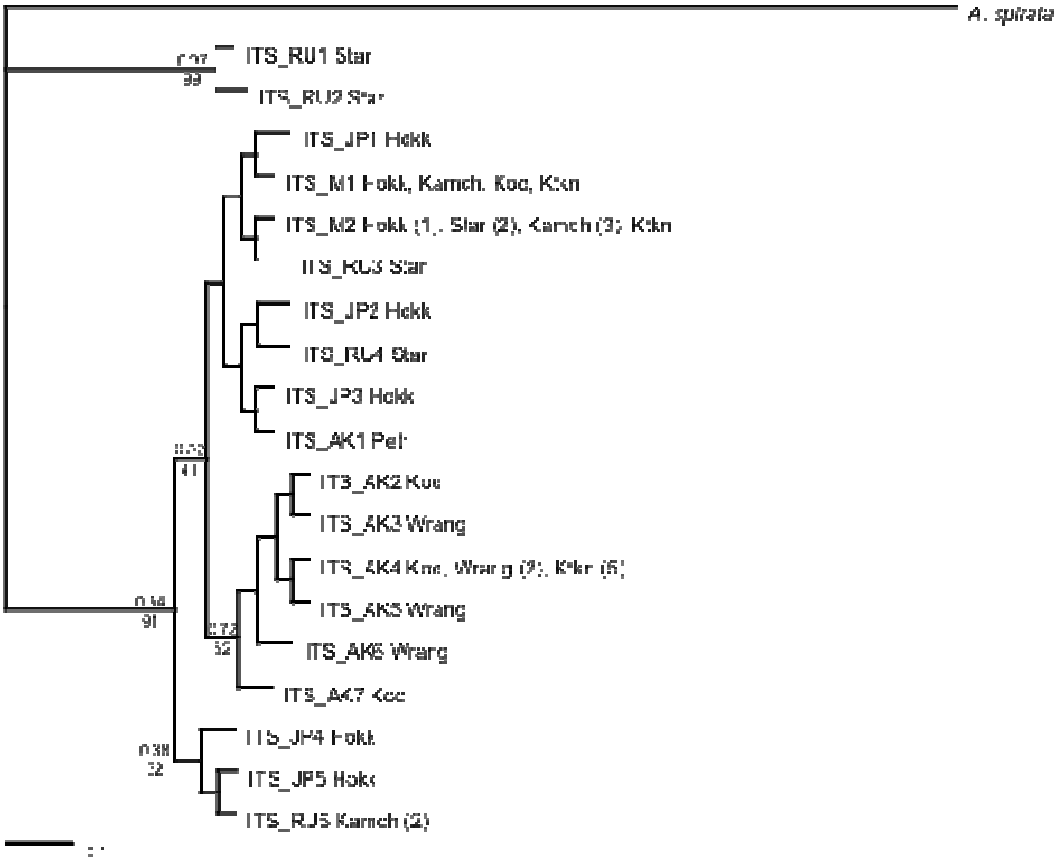
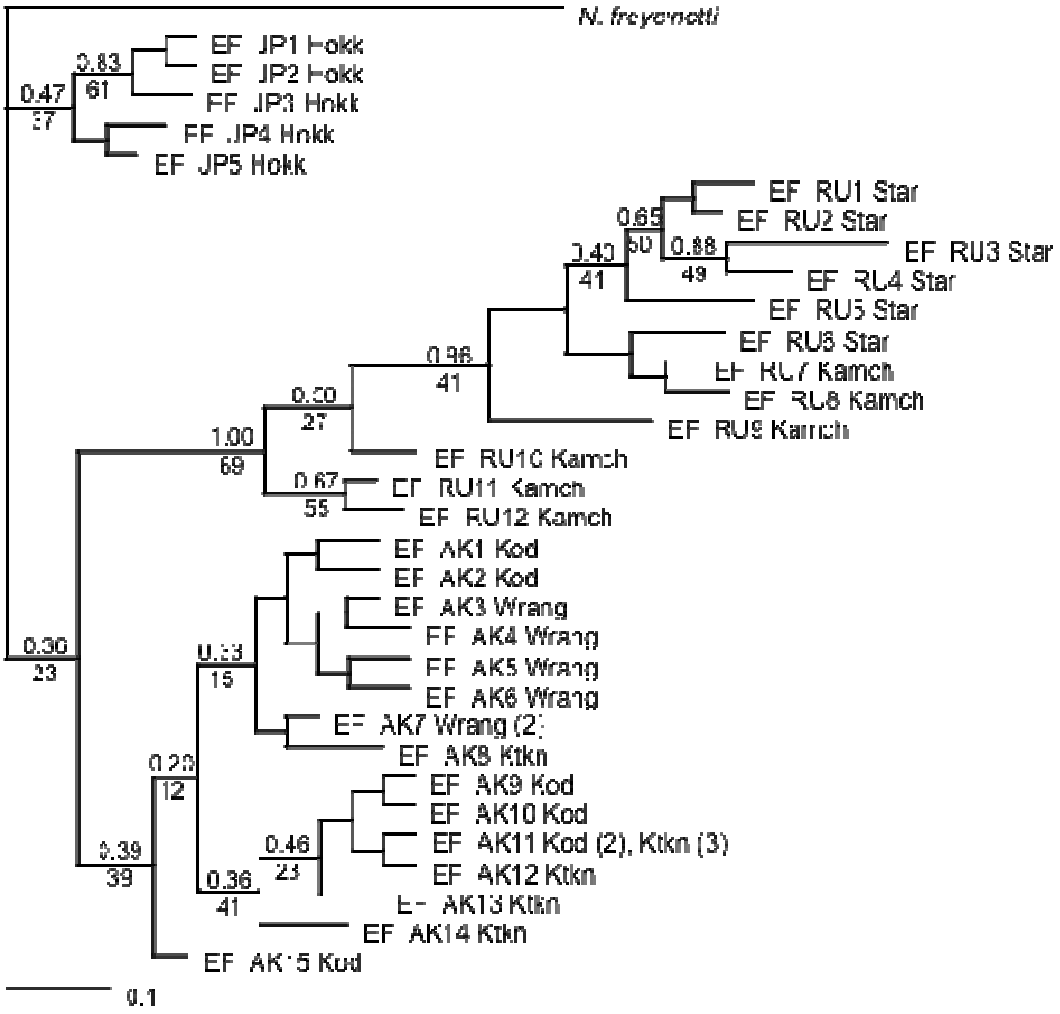

A. spirata
ITS_RU1 Star
ITS_JP1 Hokk
ITS_M1 Hokk, Kamch, Koc, Kkn
ITS_JP2 Hokk
ITS_JP3 Hokk
ITS_AK1 Pet
ITS_AK2 Koc
ITS_AK3 Wrang
ITS_AK5 Wrang
ITS_AK7 Koc
ITS_JP4 Hokk
ITS_JP5 Hokk
ITS_RU5 Kamch (2)

Figure 3c


N. freyeneti
EF JP1 Hokk
EF JP2 Hokk
EF JP3 Hokk
EF JP4 Hokk
EF JP5 Hokk
0.83
61
0.47
27
EF RU1 Star
EF RU2 Star
EF RU3 Star
EF RU4 Star
EF RU5 Star
EF RU6 Star
EF RU7 Kamch
EF RU8 Kamch
EF RU9 Kamch
EF RU10 Kamch
EF RU11 Kamch
EF RU12 Kamch
0.65
50
0.88
49
0.43
41
0.96
41
0.50
27
1.00
69
0.67
55
EF AK1 Kod
EF AK2 Kod
EF AK3 Wrang
EF AK4 Wrang
EF AK5 Wrang
EF AK6 Wrang
EF AK7 Wrang (2)
EF AK8 Ktkn
EF AK9 Kod
EF AK10 Kod
EF AK11 Kod (2), Ktkn (3)
EF AK12 Ktkn
EF AK13 Ktkn
EF AK14 Ktkn
EF AK15 Kod
0.30
23
0.23
15
0.23
12
0.39
39
0.46
23
0.36
41
0.1

The phylogenetic relationships of the EF1a gene were more similar to COI in that most haplotypes were sorted into eastern and western clades (Figure 3C). There were three major lineages of *N. lima,* and these lineages were not monophyletic with respect to the outgroup. All Alaskan samples formed a monophyletic clade but was not well-supported (posterior probability = 0.39, bootstrap percentage = 39). Hokkaido samples also formed a monophyletic clade that was not well supported (posterior probability = 0.47, bootstrap percentage = 37) even though there were better supported lineages within this clade. The best-supported clade for EF1a was a lineage comprising all of the Kamchatka and Sakhalin Island samples (posterior probability = 0.69, bootstrap percentage = 100). These populations were not reciprocally monophyletic, but structure between them was evident as all but one of the Sakhalin samples formed a clade nested within the Kamchatka samples.

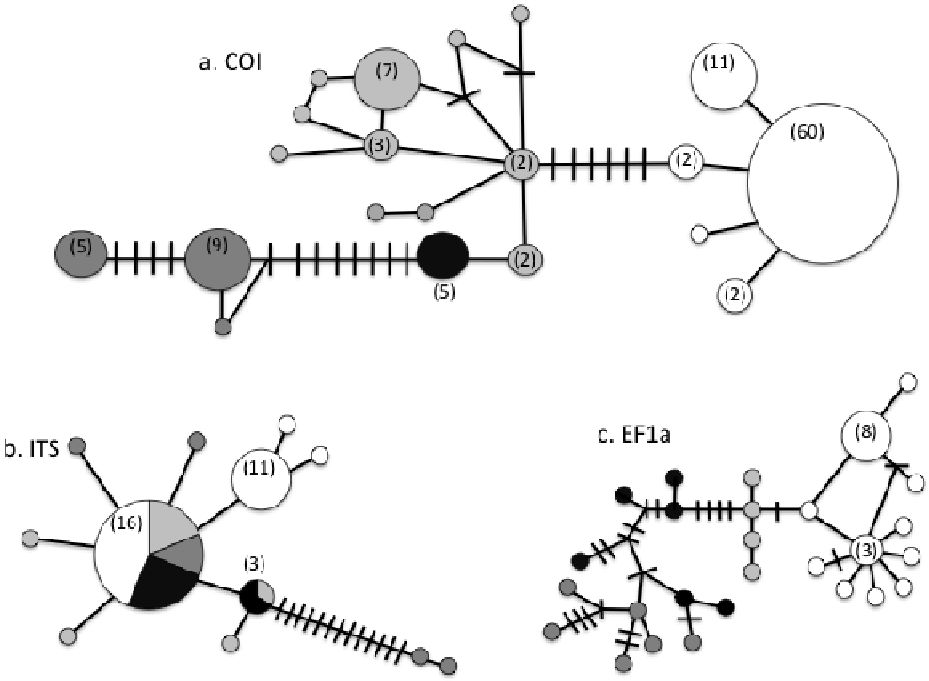
The TCS parsimony analysis produced a single network for *N. lima* for each locus studied although I had to use a network connection limit of 90% for ITS. For COI, there were 20 haplotypes sampled from 119 individuals (Figure 4a). The parsimony network produced a similar topology to the Bayesian analysis with no haplotypes shared between eastern and western regions with the exception of the Adak samples. On Sakhalin Island, there were only two common but divergent haplotypes differing by 5 bp that were mostly separated geographically between Starodubskoye and Kholmsk, but were shared between the two localities (Figure 3a and 4a).

As in the cladistic and Bayesian analyses, ITS showed little geographic structure and more admixture between populations while EF1a showed geographical structure and no mixing of haplotypes within populations (Figure 4b and 4c). EF1a had a common, central haplotype that was shared across the entire distribution with rare haplotypes branching off. ITS did not have a centralized haplotype and was lacking any haplotype occurring frequently. The most abundant haplotype was found only in Alaska and was found at multiple localities. In each case, there were rare haplotypes that were fairly divergent, all of which were from the western Pacific. Specifically, in the ITS network there were two divergent haplotypes that had 13 substitutions separating them from the other haplotypes. In the EF1a network, there were many divergent haplotypes with several substitutions separating them from others.

Population structure

Pairwise Φ_{ST} estimates for COI showed strong spatial patterns (Table 2). Comparisons between the eastern and western Pacific sites were all statistically significant, ranging from 0.826 to 1.00. Within each region, Φ_{ST}s were more variable: in the west, all pairwise estimates were significant and ranged from 0.436 to 0.930 and in the east, only a third of the estimates were significant and ranged from -0.010 to 0.714. Although eastern and western estimates overlapped, the mean pairwise Φ_{ST} was lower in the east (0.196) than the west (0.764). The smallest Φ_{ST} value in the western Pacific was found between Hokkaido and Kamchatka. In the eastern Pacific, almost all estimates of Φ_{ST}

Figure 4. Parsimony networks constructed in TCS v1.18 (Clement *et al. 2*000. Numbers indicate number of individuals if greater than one. Proportion of individuals from each region are indicated by color as follows: White- Alaska, light gray- Hokkaido, dark gray- Sakhalin Island, black- Kamchatka. a) COI b) ITS c) EF1a.



were less than 0.084 and not significant. The only exceptions were the comparisons involving Wrangell, which ranged in differentiation from 0.542 to 0.714 from the other eastern localities and were all highly significant. The population sampled at Wrangell was highly divergent because 11 individuals possessed a haplotype endemic to this site (Figure 3). The high frequency of this unique haplotype in Wrangell pushed the Φ-statistics upward for all population comparisons with this sample locality, although these estimates were not as high as the estimates between western populations.

For the full dataset, ITS showed significant Φ_{ST}s between eastern and western populations but not between populations within either region. Pairwise estimates ranged from 0.175-0.452 across the Pacific and were larger than all estimates within the eastern or western Pacific. Despite an absence of significance for both regions, Φ-statistics were larger within the west than within the east except for Hokkaido/Kamchatka, which had a Φ_{ST} of -0.099. When recombinant sites were removed, ITS showed similar pairwise results with higher Φ-statistics between eastern and western comparisons and within the west, although none were significant (results not shown).

For EF1a, the full dataset showed the same patterns as ITS with a few exceptions. Φ-statistics in the western Pacific were significant for all population comparisons and were also significant in the eastern Pacific for all comparisons with Wrangell. Φ-statistics were largest across the Pacific, ranging from 0.618-0.740. Also, comparisons in the west generally showed larger genetic differentiation than in the east except Sakhalin and Kamchatka, which had a

Table 2. COI Pairwise Φ_{ST} comparisons for each sampling station for *N.lima* below diagonal. *P*-values above diagonal. Adak was excluded because of the low sample size. Shaded areas indicate comparisons between the east and west Pacific regions. * indicates $P < 0.05$; **indicates $P<0.01$

Region		Hokk	Kho	Star	Kamch	Kod	Hom	Sew	Jun	Wrang	Ktkn
West Pacific	Hokk	-	0.0000	0.0000	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Kho	0.843**	-	0.0169	0.0075	0.0002	0.0010	0.0002	0.0003	0.0000	0.0002
	Star	0.847**	0.602*	-	0.0012	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Kamch	0.436**	0.925**	0.930*	-	0.0008	0.0025	0.0006	0.0004	0.0001	0.0003
East Pacific	Kod	0.853**	0.966**	0.962**	1.00**	-	0.1636	0.9999	0.9999	0.0003	0.4853
	Hom	0.826**	0.936**	0.939**	0.966**	0.042	-	0.1103	0.1476	0.0006	0.5239
	Sew	0.867**	0.974**	0.968**	1.00**	0.000	0.084	-	0.9999	0.0000	0.1064
	Jun	0.857**	0.969**	0.964**	1.00**	0.000	0.053	0.000	-	0.0001	0.4820
	Wrang	0.861**	0.953**	0.951**	0.970**	0.680**	0.542**	0.714**	0.689**	-	0.0000
	Ktkn	0.849**	0.950**	0.949**	0.969**	0.029	-0.010	0.057	0.037	0.564**	-

lower differentiation value of 0.253. With recombinant sites removed, EF1a results were more variable and did not show strong spatial patterns (results not shown).

In SAMOVA, each locus produced a different result. For the mitochondrial locus, SAMOVA found that three groups maximized the among-group variance while minimizing the variance within groups (Table 3). Group one contained Hokkaido and Kamchatka, group two was composed of Starodubskoye, and group three contained all of the eastern Pacific sites: Kodiak, Wrangell, and Ketchikan. The two nuclear loci showed different groupings for all values of k but were fairly consistent with each other with respect to eastern and western groups. However, the number of groups that maximized among-group variance was not consistent between ITS and EF1a. Estimates from SAMOVA suggested the ITS data best supported only two panmictic groups: Starodubskoye in group one and the rest of the localities in group two. The EF1a data supported four panmictic groups, with each of the western sites comprising a different group and all of the eastern sites combined for group four. In summary, two out of three loci grouped Hokkaido and Kamchatka together, all loci separated Starodubskoye into its own group, and Wrangell was distinct only for EF1a but was also separated when k=4 for COI although this grouping was not optimal (Φ_{CT}=0.847).

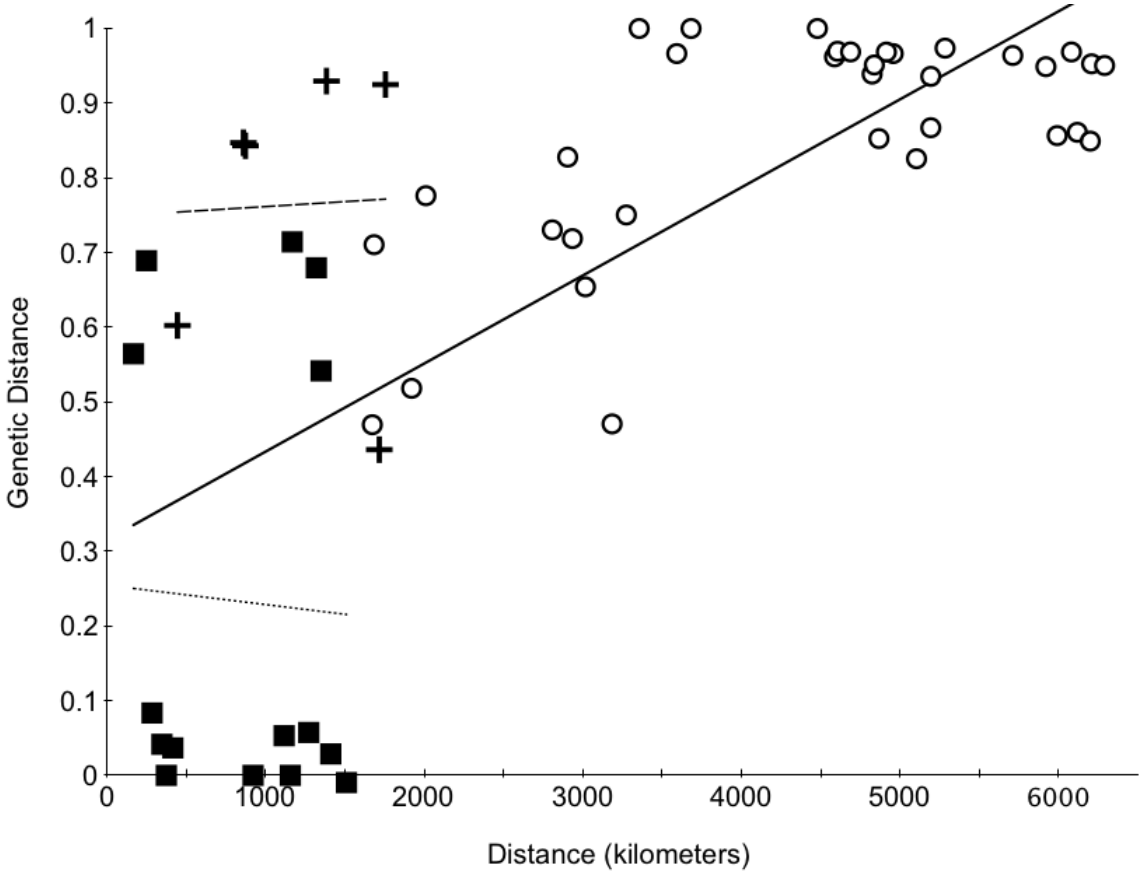
The Mantel test for isolation by distance (IBD) of COI was significant when all samples were included (R^2= 0.487, p=0.001; Figure 5). However, when eastern and western regions were analyzed separately, significant isolation by distance was not detected (eastern Pacific, R^2 was 0.0019 and p=0.623; western

Table 3. Summary of SAMOVA. k specifies the number of groups hypothesized .

Locus	*k*	Groups	Φ_{CT}	Φ_{SC}	Φ_{ST}
COI	2	[Hokk, Star, Kamch] [Kod, Wrang, Ktkn]	0.559	0.798	0.911
	3	**[Hokk, Kamch] [Star] [Kod, Wrang, Ktkn]**	**0.863**	**0.364**	**0.913**
	4	[Hokk, Kamch] [Star] [Kod, Ktkn] [Wrang]	0.847	0.328	0.897
	5	[Hokk, Kamch] [Star] [Kod] [Wrang] [Ktkn]	0.730	0.594	0.890
ITS	**2**	**[Hokk, Kamch, Kod, Wrang, Ktkn] [Star]**	**0.449**	**-0.063**	**0.414**
	3	[Hokk, Kod, Wrang, Ktkn] [Star] [Kamch]	0.324	0.272	-0.077
	4	[Hokk] [Star] [Kamch] [Kod, Wrang, Ktkn]	0.303	-0.137	0.208
	5	[Hokk] [Star] [Kamch] [Kod] [Wrang, Ktkn]	0.293	-0.169	0.174
EF1a	2	[Hokk, Kamch, Kod, Wrang, Ktkn] [Star]	0.387	0.330	0.589
	3	[Hokk, Kod, Wrang, Ktkn] [Star] [Kamch]	0.388	0.216	0.520
	4	**[Hokk, Kamch] [Star] [Kod, Ktkn] [Wrang]**	**0.416**	**0.075**	**0.459**
	5	[Hokk, Kamch] [Star] [Kod] [Wrang] [Ktkn]	0.365	0.114	0.438

Boldface groups are the ones that maximize among group variance and minimize within group variance

Figure 5. Isolation by distance patterns with genetic distance expressed as pairwise ϕ_{ST}. Details are shown for within Eastern Pacific (closed squares), within the western Pacific (crosses) as well as for all combined samples (open circles). Trend lines are as follows: solid line- all samples, long dashes- western Pacific, short dashes- eastern Pacific. Only the slope of the regression for all samples (solid line) is significantly different from 0.



Pacific, R^2=0.0012 and p=0.58). The Mantel test for the full dataset of ITS was significant across the six sites (R^2= 0.779, p=0.044) but EF1a was not (R^2= 0.600, p=0.066). Because there were only three localities for each nuclear locus within the eastern and western Pacific regions, the IBD estimates for these loci are not informative and, therefore, are not provided for within-region analyses.

Population divergence models

There were three population pairwise comparisons in IMA between eastern Pacific populations: Kodiak/Wrangell, Wrangell/Ketchikan, and Kodiak/Ketchikan. All models with no migration were rejected for all three east vs. east comparisons (Table 4). Overall, the divergence between eastern populations was minimal and this was reflected in the migration and divergence time estimates (Table 5). Migration rate point estimates ranged from 0 to 12.29. Posterior probabilities of the migration rate parameter was not resolved, and in five out of six estimates in the east increased infinitely towards zero with no clear peaks in likelihood (Figure 6). Theta estimates were low in the eastern populations, 0.0091-0.53, which translates to effective population sizes of 8,945-60,105 individuals (Table 5). Divergence times for Wrangell/Ketchikan (t=8.93 x 10^4 years ago) and Kodiak/Ketchikan (4.16 x 10^4 years ago) were estimated to occur during the Pleistocene glacial cycles (Table 6; Figure 7a). The Kodiak/Wrangell divergence time estimate had no clear peak in likelihood, and IMA was not able to distinguish between a recent or more ancient time of divergence. The divergence time confidence intervals for Kodiak/Wrangell and Wrangell/Ketchikan were not

Table 4. Nested models test for *N. lima* pairs of populations including all three loci. The values presented correspond to two times the log-likelihood ratio between the full model given by IMa and a simpler model presented in the first column. 2LLR values in bold indicate the probability to achieve the test statistic by chance is <0.05. 2LLR values that are not in bold fail to be rejected when compared.

		East vs. East			West vs. West		
Model (Θ)	DF	Kod/Wrang	Wrang/Ktkn	Kod/Ktkn	Hokk/Star	Star/Kamch	Hokk/Kamch
$\theta_1, \theta_2, \theta_A, m_1=m_2$	1	**4.1779**	1.8496	0.1839	0.0280	0.6925	-2.4097
$\theta_1, \theta_2, \theta_A, m_1, m_2=0$	1	3.7640	0.0011	2.6768	**10.7748**	0.0001	-0.0023
$\theta_1, \theta_2, \theta_A, m_1=0, m_2$	1	**3.1880**	3.6049	2.2295	**8.5295**	1.4879	-0.0032
$\theta_1, \theta_2, \theta_A, m_1=m_2=0$	2	**16.0740**	**20.0481**	**20.0255**	**10.7751**	4.7375	-0.0033
$\theta_1=\theta_2, \theta_A, m_1, m_2$	1	2.6604	**4.9042**	2.5229	**7.1406**	1.6002	**5.3924**
$\theta_1=\theta_2=\theta_A, m_1, m_2$	2	**7.1675**	4.9025	4.8191	**9.2969**	2.9728	**7.8039**
$\theta_1=\theta_2, \theta_A, m_1=m_2$	2	**8.3082**	1.8536	5.5666	2.0216	3.4996	**7.5900**
$\theta_1=\theta_2, \theta_A, m_1=m_2=0$	3	**39.4010**	**46.0750**	**23.5238**	**11.7366**	**8.8896**	**19.4408**
$\theta_1=\theta_2=\theta_A, m_1=m_2$	3	**9.6913**	6.9947	6.1677	**10.9457**	5.4685	**10.9044**
$\theta_1=\theta_2=\theta_A, m_1=m_2=0$	4	**73.6386**	**49.3647**	**75.7686**	**12.1137**	**15.6760**	**22.0558**
$\theta_1=\theta_A, \theta_2, m_1, m_2$	1	**5.3934**	**4.9003**	2.5572	**8.5918**	0.6940	2.4292
$\theta_1=\theta_A, \theta_2, m_1=m_2$	2	**7.6947**	**6.4462**	4.5398	**10.6716**	1.6925	3.5133
$\theta_1=\theta_A, \theta_2, m_1=m_2=0$	3	**25.3080**	**20.9226**	**55.9768**	**10.7717**	**8.8742**	5.7828
$\theta_2=\theta_A, \theta_1, m_1, m_2$	1	**5.3923**	3.5955	2.1520	**8.9777**	1.4256	0.0710
$\theta_2=\theta_A, \theta_1, m_1=m_2$	2	**6.9292**	**6.4390**	4.6572	**10.9124**	3.5205	0.1607
$\theta_2=\theta_A, \theta_1, m_1=m_2=0$	3	**61.4132**	**38.3323**	**49.0229**	**11.8904**	**15.3102**	5.8991

Table 5. IMA estimates for eastern Pacific population comparisons of theta (θ), migration rate (m), and population divergence times (T; millions of years) inferred with the program IMA for *Nucella lima*. 90% highest posterior density or HPD intervals are found in parentheses. HPD intervals are stated as '?' if it was difficult to define upper and lower bounds on the posterior probability distribution.

Samples	Model	θ_1	θ_2	θ_A	m_1	m_2	T
Kod/	θ_1, θ_2, θ_A,	0.049	0.23	?	?	12.29	?
Wrang	m_1, m_2	(0.02-1.58)	(0.15-7.38)	?	?	(2.39-79.88)	?
Wrang/	θ_1, θ_2, θ_A,	0.0091	0.056	0.023	?	?	0.0893
Ktkn	m_1, m_2	(0.0039 - 0.23)	(0.023 -1.42)	?	?	?	?
Kod/	θ_1, θ_2, θ_A,	0.24	0.53	0.15	?	?	0.0416
Ktkn	m_1, m_2	(0.078-3.34)	(0.32-10.58)	(0.056-3.50)	?	?	(0.0226-2.3)

Figure 6. IMA migration rate posterior probability distributions. Black line is migration rate into the first population; gray line is migration rate into the second population. The only population comparisons included are those for which all models of no migration could be rejected in the nested models analysis in IMA.

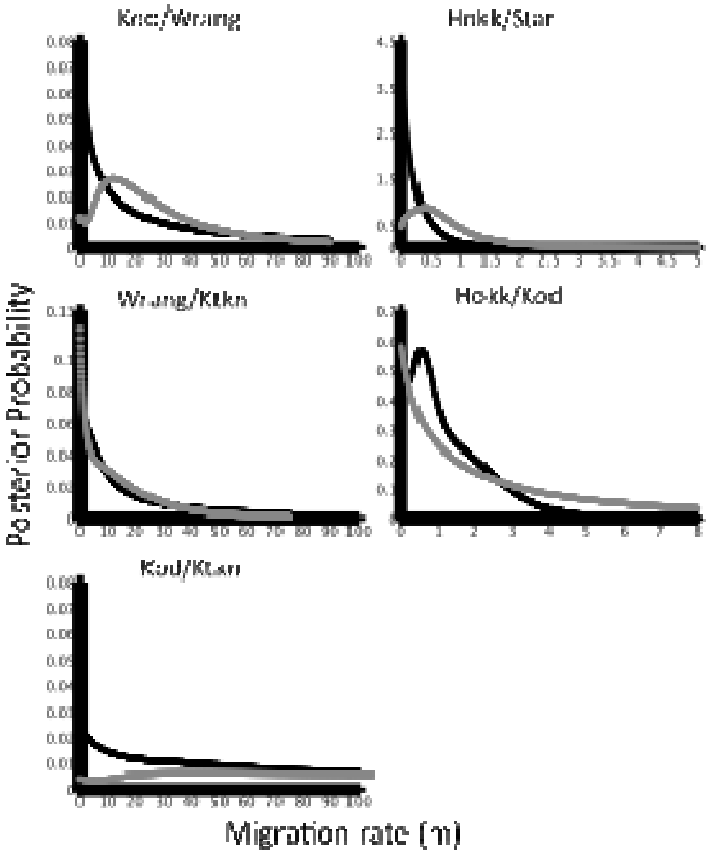


Table 6. Nested models test for *N. lima* pairs of populations including all three loci. The values presented correspond to two times the log-likelihood ratio between the full model given by IMa and a simpler model presented in the first column. 2LLR values in bold indicate the probability to achieve the test statistic by chance is <0.05. 2LLR values that are not in bold fail to be rejected when compared.

		East vs. West								
Model (Θ)	DF	Hokk/ Kod	Hokk/ Wrang	Hokk/ Ktkn	Star/ Kod	Star/ Wrang	Star/ Ktkn	Kamch/ Kod	Kamch/ Wrang	Kamch/ Ktkn
$\theta_1; \theta_2; \theta_A; m_1=m_2$	1	**5.1470**	3.4479	**4.7834**	1.0963	1.2138	1.4295	-0.0006	3.3301	-0.0006
$\theta_1; \theta_2; \theta_A; m_1; m_2=0$	1	0.0047	3.4474	3.5124	0.0881	0.2617	0.2144	-0.0016	1.8647	-0.0016
$\theta_1; \theta_2; \theta_A; m_1=0; m_2$	1	3.4168	0.1262	**5.2291**	**4.0191**	2.4339	2.7774	0.0026	0.0018	0.0026
$\theta_1; \theta_2; \theta_A; m_1=m_2=0$	2	**9.2006**	3.4455	5.2257	5.9135	2.4326	3.1726	-0.0018	3.5345	-0.0018
$\theta_1=\theta_2; \theta_A; m_1; m_2$	1	**12.7625**	10.8573	**12.0813**	**7.1688**	**4.5425**	**4.7158**	1.7364	2.2968	1.7364
$\theta_1=\theta_2=\theta_A; m_1; m_2$	2	**12.7906**	**11.4446**	**12.9612**	**7.5171**	**6.1035**	**6.2577**	**6.7555**	2.5074	**6.7555**
$\theta_1=\theta_2; \theta_A; m_1=m_2$	2	**15.5961**	**13.5601**	**12.0794**	**8.9391**	4.5713	**7.7684**	1.7355	3.8655	1.7355
$\theta_1=\theta_2; \theta_A; m_1=m_2=0$	3	**40.4056**	**13.5588**	**12.0793**	**15.9010**	**15.2315**	**8.9584**	1.7355	5.5930	1.7355
$\theta_1=\theta_2=\theta_A; m_1=m_2$	3	**15.7676**	**13.9652**	**12.9611**	**10.8336**	7.6408	**9.2450**	7.0067	3.9729	7.0067
$\theta_1=\theta_2=\theta_A; m_1=m_2=0$	4	**42.4056**	**14.5186**	**12.9611**	**19.4778**	**21.0259**	**38.2580**	**10.9942**	6.6885	**10.9942**
$\theta_1=\theta_A; \theta_2; m_1; m_2$	1	3.2957	**4.1540**	**5.9347**	1.3439	0.4819	0.8655	**6.6270**	2.4257	**6.6270**
$\theta_1=\theta_A; \theta_2; m_1=m_2$	2	5.2602	4.0539	**6.2036**	2.1465	1.4618	2.4959	**6.9962**	3.8544	**6.9962**
$\theta_1=\theta_A; \theta_2; m_1=m_2=0$	3	**10.8314**	4.0536	6.2035	5.9766	3.0554	**13.7153**	**9.5353**	4.9599	**9.5353**
$\theta_2=\theta_A; \theta_1; m_1; m_2$	1	0.0087	**4.1383**	**5.2503**	0.5901	1.1476	1.3802	**6.6415**	2.4378	**6.6415**
$\theta_2=\theta_A; \theta_1; m_1=m_2$	2	5.3900	5.8291	**7.0986**	1.1152	2.3995	3.2490	**7.0054**	3.8343	**7.0054**
$\theta_2=\theta_A; \theta_1; m_1=m_2=0$	3	**20.5897**	**10.4856**	7.7903	**11.2903**	**14.2274**	**34.7196**	**9.8163**	4.6751	**9.8163**

provided because the upper bounds on the posterior probability distribution could not be defined, and therefore, these point estimates were unreliable (Figure 7a).

There were also three population comparisons in IMA between western Pacific populations: Hokkaido/Starodubskoye, Starodubskoye/Kamchatka, and Hokkaido/Kamchatka. None of the models with no migration could be rejected for Starodubskoye/Kamchatka or Hokkaido/Kamchatka (Table 7). Although a no-migration model was rejected for Hokkaido/Starodubskoye, migration rate estimates between these sites were low, 0 and 0.39 (Table 7; Figure 6). Theta estimates ranged from 1.05 to 11.48, which translated to effective population size estimates of 103,876-1,134,165 individuals (Table 7). Starodubskoye/Kamchatka and Hokkaido/Kamchatka had dates of divergence pre-dating two out of the three divergences in the eastern Pacific but still occurring within the Pleistocene climate cycles, 5.34×10^5 and 1.39×10^5 years ago, respectively (Table 7). The divergence estimates between Hokkaido/Starodubskoye were unreliable because there was no clear peak in the posterior probability and the point estimate was located outside of the 90% confidence interval (Figure 7b).

There were nine population comparisons in IMA between eastern and western Pacific populations: Hokkaido/Kodiak, Hokkaido/Wrangell, Hokkaido/Ketchikan, Starodubskoye/Kodiak, Starodubskoye/Wrangell, Starodubskoye/Ketchikan, Kamchatka/Kodiak, Kamchatka/Wrangell, and Kamchatka/Ketchikan. All models with no migration could not be rejected for all population comparisons except Hokkaido/Kodiak (Table 6). Despite rejecting a no-migration model, migration rates for Hokkaido/Kodiak were unreliable and

Figure 7. IMA divergence time posterior probability distributions for comparisons among a) eastern population, b) western populations, and c) eastern vs. western populations.

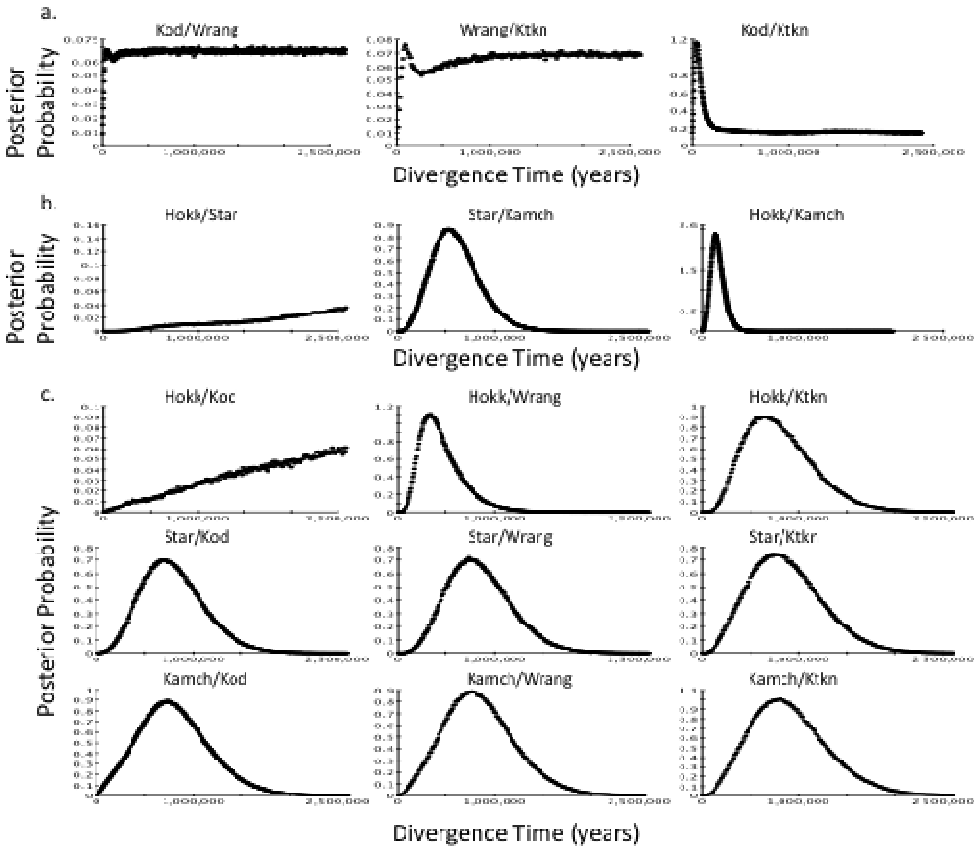


Table 7. IMA estimates for western Pacific population comparisons of theta (θ), migration rate (m), and population divergence times (T; millions of years) inferred with the program IMA for *Nucella lima*. 90% highest posterior density or HPD intervals are found in parentheses. HPD intervals are stated as '?' if it was difficult to define upper and lower bounds on the posterior probability distribution.

Samples	Model	θ_1	θ_2	θ_A	m_1	m_2	T
Hokk/	θ_1, θ_2, θ_A,	4.51	2.94	8.01	?	0.39	?
Star	m_1, m_2	(2.43-9.73)	(1.39-9.52)	(5.85-418.19)	?	(0.048-2.87)	?
Star/	θ_1, θ_2, θ_A,	3.10	1.14	6.77	-	-	0.534
Kamch	$m_1 = m_2 = 0$	(1.56-12.29)	(0.46-4.79)	(3.35-42.57)	-	-	(0.207 – 1.17)
Hokk/	θ_1, θ_2, θ_A,	11.48	1.05	1.83	-	-	0.139
Kamch	$m_1 = m_2 = 0$	(5.69-167.26)	(0.43-6.91)	(0.78-5.63)	-	-	(0.0484 -0.329)

produced no clear peak in likelihood and increased in likelihood infinitely towards zero (Figure 6). Theta estimates for all nine comparisons ranged from 0.22 to 5.98 or an effective population size range from 38,770-1,147,270 individuals (Table 8). The divergence times between eastern and western Pacific populations overlapped with those in the western Pacific but were on average higher than divergences between populations within either the eastern or western Pacific, ranging from $3.36 \times 10^5 - 8.11 \times 10^5$ years ago (Table 8). Hokkaido/Kodiak divergence estimates were unreliable because the confidence interval had no definable upper bound or clear peak in likelihood (Figure 7c). All reliable divergence time estimates between eastern and western populations coincided with the Pleistocene climate cycles.

Population size change

Because IMA analyses showed evidence for limited gene flow in all population comparisons, I estimated changes in effective population size for all six populations for which I had multi-locus data using BEAST. There was minimal recent growth on all six of the skyline plots, and, according to the marginal posterior distribution of the growth rate in the EBSP, constant population size could not be rejected for all six populations because a growth rate of zero was included in the 95% confidence intervals. According to the Bayes factors, in four out of six populations a constant size model was the best fit model or as equally likely as one of the other models (Table 9). It is notable that all of the western populations fit a constant size model and only one of the eastern populations did,

Table 8. IMA estimates for population comparisons between eastern and western populations of theta (θ), migration rate (m), and population divergence times (T; millions of years) inferred with the program IMA for *Nucella lima*. 90% highest posterior density or HPD intervals are found in parentheses. HPD intervals are stated as '?' if it was difficult to define upper and lower bounds on the posterior probability distribution.

Samples	Model	θ_1	θ_2	θ_A	m_1	m_2	T
Hokk/ Kod	θ_1, θ_2, θ_A, m_1, m_2	4.67 (2.80-36.09)	0.22 (0.08-1.61)	3.37 (2.23-111.14)	? ?	? ?	? ?
Hokk/ Wrang	θ_1, θ_2, θ_A, $m_1 = m_2 = 0$	4.47 (2.37-15.55)	0.61 (0.24-2.06)	2.64 (1.12-11.24)	- -	- -	0.336 (0.154-1.02)
Hokk/ Ktkn	θ_1, θ_2, θ_A, $m_1 = m_2 = 0$	5.98 (3.10-17.84)	0.82 (0.35-2.32)	2.47 (0.51-10.69)	- -	- -	0.662 (0.270-1.62)
Star/ Kod	θ_1, θ_2, θ_A, $m_1 = m_2 = 0$	3.93 (2.04-14.80)	0.63 (0.23-2.38)	7.14 (3.40-49.29)	- -	- -	0.690 (0.268-0.153)
Star/ Wrang	θ_1, θ_2, θ_A, $m_1 = m_2 = 0$	3.92 (2.04-14.3)	0.787 (0.31-2.51)	7.13 (3.38-48.42)	- -	- -	0.750 (0.340-1.60)
Star/ Ktkn	θ_1, θ_2, θ_A, $m_1 = m_2 = 0$	3.82 (1.93-13.52)	0.92 (0.39-3.06)	6.49 (3.44-40.11)	- -	- -	0.750 (0.257-1.74)
Kamch/ Kod	θ_1, θ_2, θ_A, $m_1 = m_2 = 0$	0.89 (0.34-7.22)	0.68 (0.25-3.39)	? ?	- -	- -	0.748 (0.181 -1.66)
Kamch/ Wrang	θ_1, θ_2, θ_A, $m_1 = m_2 = 0$	0.89 (0.34-6.16)	1.01 (0.42-3.21)	? ?	- -	- -	0.746 (0.233-1.70)
Kamch/ Ktkn	θ_1, θ_2, θ_A, $m_1 = m_2 = 0$	0.83 (0.31-4.29)	0.77 (0.30-2.22)	? ?	- -	- -	0.811 (0.270-0.186)

Table 9. Bayes Factor Tests comparing demographic models for 6 populations of *N. lima*. Bayes Factors correspond to row by column comparisons. Boldface corresponds to the best fitting demographic model

Population	Model	Ln P(Model)	SE	Bayes Factors		
				Constant	Exponential	EBSP
Hokk	**Constant**	-2198.446	0.06	-	4.429*	4.016*
	Exponential	-2199.934	0.067	0.226	-	0.907
	EBSP	-2199.837	0.066	0.249	1.103	-
Star	Constant	-2187.869	0.052	-	0.655	172048.9***
	Exponential	-2199.924	0.008	1.528	-	262863***
	EBSP	-2187.445	0.046	0	0	-
Kamch	**Constant**	-1367.875	0.034	-	5.707*	1.513
	Exponential	-1369.616	0.036	0.175	-	0.265
	EBSP	-1368.289	0.037	0.661	3.771*	-
Kod	**Constant**	-1359.951	0.031	-	3.378*	1.013
	Exponential	-1361.168	0.035	0.296	-	0.3
	EBSP	-1359.964	0.033	0.987	3.334*	-
Wrang	Constant	-2056.611	0.038	-	0	3.331*
	Exponential	-2043.76	0.039	381275.2***	-	1270027***
	EBSP	-2057.814	0.034	0.3	0	-
Ktkn	Constant	-1367.509	0.054	-	0.487	1.047
	Exponential	-1366.79	0.041	2.053	-	2.149
	EBSP	-1367.555	0.052	0.955	0.465	-

Strength of Bayes Factor evidence based on Jeffreys (1961): *=substantial; **=strong; ***=decisive

while the other two populations, Wrangell and Ketchikan in the east, show evidence for an exponential growth rate model. Overall, using the two methods of evaluating population growth in BEAST, none of the populations analyzed here show evidence for an increase in population size after the LGM.

Discussion

Trans-Pacific divergence and timing

Substantial east-west differentiation of *N. lima* across the Pacific Ocean is consistent with the timing of the glacial cycles and may be a direct result of isolation in refugia during glacial periods. There is substantial and highly significant differentiation between western and eastern samples ($\Phi_{ST} > 0.8$) at the mtDNA locus. At all loci, the Φ-statistics between populations across the Pacific Ocean are greater than the Φ-statistics among populations within either the eastern or western Pacific. Using a genus-specific molecular clock based on the substitution rate, minimal east-west divergence times are in excess of 150 kyr, which precedes the end of the LGM 20 kyr. This pattern of differentiation between the east and west, dating before the LGM, provides evidence for refugia on both sides of the Pacific.

Comparatively, congeneric species, *Nucella lamellosa* and *N. ostrina*, which have a similar dispersal capability show a hierarchical Φ-statistic value between a northern and southern grouping of populations in the eastern Pacific of 0.108 and 0, respectively (Marko 2004). Φ-statistics in *N. lima* are higher across the ocean as well as within regions than detected in these two congeneric

species. Compared to other species in this region showing divergence between eastern and western populations, Φ-statistics in *N. lima* are abnormally large. In Pacific cod, mtDNA Φ-statistics across the Pacific are significant, but not greater than 0.30, which could be explained by the large dispersal capability of this fish (Canino *et al.* 2010). In a marine invertebrate, the lined shore crab, *Pachygrapsus crassipes*, showed a large mtDNA Φ_{CT} of 0.84 between the population in Korea and populations in the eastern Pacific and using sequence divergence dated this split to 0.8-1.2 MYA, which is expected to predate the actual timing of the population divergence (Cassone & Boulding 2006). Although this estimate of trans-Pacific divergence is congruent with *N. lima*, *P. crassipes* does not have a continuous distribution across the North Pacific Ocean and despite the long dispersal capability, this high differentiation value is not unexpected given the long geographic distance between the populations.

It is this large genetic differentiation between the east and west and the coincident geographic distance that creates a pattern of isolation by distance when all samples are included. When only samples within a region are tested, which includes over 1500 km of coastline, there is no IBD pattern indicating the large Φ-statistics across the Pacific Ocean are driving the significant IBD pattern when all samples are included in the regression. This may be the effect of a strong geographic barrier to gene flow as opposed to an IBD pattern due to migration-drift equilibrium alone. When the regressions within each region are compared, it is evident that the western samples have larger Φ-statistics between sites than in the eastern Pacific even though the geographic distances are

overlapping.

Species may have resisted isolation and potentially speciation by recolonizing glaciated areas during interglacials and maintaining gene flow among persistent populations due to the lack of long-term geographical barriers. *Nucella lima* was likely incapable of recolonizing the northern latitudes in the relatively short interglacial periods due to the absence of a planktonic or widely dispersing life history stage and, thus, did not maintain gene flow between previously separated populations. The IMA nested models analysis confirms a lack of gene flow for all population comparisons except for Hokkaido/Kodiak, but IMA estimates gene flow for this population comparison to be limited.

The nuclear genes support similar phylogenetic relationships as the mitochondrial gene, even though the clades are not as strongly structured geographically. The eastern and western clades are not reciprocally monophyletic at all loci, although the deep divergences across the Pacific are evident in the gene trees. For COI and EF1a, there is strong support for geographically restricted clades relating not only to the eastern and western Pacific but also within the western Pacific although these clades are not consistent across the two genes. The western samples may contain multiple lineages in the process of speciation, although without further evidence of reproductive isolation or continued limited gene flow, this cannot be determined with any certainty.

There are many documented reasons for discordant mitochondrial and nuclear gene trees (reviewed by Buckley *et al.* 2006). Among the reasons are

incomplete lineage sorting, genetic polymorphism, hybridization, and introgression. Incomplete lineage sorting could be the contributing factor because many clades are restricted geographically, although monophyly has not been achieved in all populations. Hybridization and introgression, however, cannot be ruled out without more evidence of reproductive isolation because the range of *N. lima* does partially overlap with congeneric species in both the east (*N. canaliculata, N. lamellosa*, and *N. ostrina*) and west (*N. freycinetti* and *N. heyseana*). This advanced lineage sorting in the nuclear genes indicate some combination of low migration rates, small population sizes and deep divergence times and given the demographic estimates from IMA, it looks to be primarily a combination of low migration and deep divergences.

Molecular diversity and structure in the Eastern Pacific

Multilocus genetic diversity in *N. lima* is lower in the eastern Pacific than in the western Pacific; diversity of eastern samples of *N. lima* is also low compared to other intertidal invertebrates in the same region (Marko *et al.* 2010), with many eastern sites of *N. lima* having nucleotide and haplotype diversities of zero. Lower diversity in the east does not reflect a founder effect associated with eastern colonization from the west after the LGM for the following reasons: because there are deep divergences between the east and west, the haplotypes in the east do not represent a subset of the haplotypes found in the west, and there is not a gradient of decreased diversity from west to east.

Low genetic diversity in the east is more likely a result of expansion and

recolonization from an eastern refuge, such as Wrangell, after the LGM even though there was minimal evidence supporting a population expansion. This could be due to a lack of information in the three loci used rather than evidence that population size has remained constant over time. Because both haplotype and nucleotide diversities are low, effective population sizes are likely small and there are not enough coalescences in the data for BEAST to estimate a population expansion even though it is probable that one occurred. Because Wrangell contains a frequent, unique haplotype and contains a relatively high diversity in the eastern Pacific, it is likely a glacial refuge for *N. lima*. The distribution of plants and animals in the Alexander Archipelago suggest several ice-free regions that may have served as glacial refugia in at least the LGM (Carrara *et al.* 2007). Although, Wrangell was not specifically included in a putative ice-free location, areas nearby could have harbored *N. lima* and other marine intertidal species and produced this genetic pattern.

Other than Wrangell, no other sites showed significant Φ-statistics in the pairwise analysis. While this may indicate sufficient gene flow to prevent differentiation and IMA rejected a no-migration model for all populations in the east, the estimates of gene flow in IMA were low. The divergence times between eastern populations were estimated to be within the Pleistocene in IMA. While these times are more recent than those in the west or those between eastern and western populations, they still pre-date the LGM, which was expected to be when most speciation events should have occurred due to the most intense glacial cycles of the Pleistocene (Rand 1948).

Molecular diversity and structure in the Western Pacific

The higher Φ-statistics among sampling localities in the west suggest potential ice-age isolation among populations in the west. The western Pacific Ocean was fragmented into isolated habitats during the glacial cycles of the Pleistocene in which the Japan Sea, and possibly the Sea of Okhotsk, was largely isolated (Briggs 1974) providing multiple locations appropriate for glacial refugia. Unlike the eastern Pacific, every site sampled in the west had significant Φ-statistics from each other for both COI and EF1a and were quantitatively larger in the west for ITS but are not statistically significant from zero. Genetic pairwise Φ_{ST} estimates were high, in some cases upwards of 0.9 for COI. Among populations in the west for both COI and ITS, Kamchatka and Hokkaido show the lowest pairwise Φ-statistics but there were no COI or EF1a haplotypes shared between these sites and therefore it is unlikely they are connected strongly by gene flow. Divergence between the two localities on Sakhalin Island (Starodubskoye and Kholmsk) was considerable for COI. Studies of other taxa show evidence of separation between the Sea of Japan and Okhotsk Sea even in highly dispersing species (Canino *et al.* 2010), potentially due to the narrowing and sea level changes in the northern entrance to the Sea of Japan (Kitamura *et al*. 2001). This idea of ice age isolation in the west is also supported by the divergence time estimates in IMA. The time of divergence estimates are older than those in the eastern Pacific, occurring more than 139 kyr during the Pleistocene.

In contrast to the eastern Pacific populations, I was able to assume no migration in IMA between pairs of western Pacific populations, with the exception of the Hokkaido/Starodubskoye comparison, which still showed limited migration. It is likely that these populations were isolated during the early glacial cycles and that lack of gene flow due to the low dispersal capability of this species has maintained isolation.

More evidence needed to identify a glacial refuge is higher diversity and the presence of unique alleles. Nucleotide and haplotype diversities in the western Pacific were notably greater across all sites except Kamchatka, which only contained a single COI haplotype. The western samples in the COI haplotype network contained more less-frequent haplotypes with the conspicuous absence of a common haplotype for the entire region. Hokkaido contained the largest haplotype diversity, with many rare haplotypes. Overall, geographical restriction of haplotypes to individual sampling localities was notably higher in the west for COI than in the eastern Pacific. EF1a also showed some geographical structuring in the west indicating population isolation has been occurring for a considerable period of time in order for this nuclear gene to accumulate frequency differences among populations in this region.

Similar genetic structure is common in this area dating to the early to mid-Pleistocene in other species such as the ice goby (Kokita & Nohara 2010), hair crab (Azuma *et al.* 2007), and Pacific cod (Canino *et al*. 2010), which provides further evidence supporting multiple refugia in the western Pacific Ocean. In Pacific cod, *Gadus macrocephalus*, microsatellite data showed divergences in

the northwestern Pacific exceeded those in the eastern Pacific indicating longer isolation of populations in the west, although the Φ-statistics did not exceed 0.1 and the pattern did not appear in the mitochondrial data (Canino *et al.* 2010). *Erimacrus isenbeckii,* the hair crab also showed weak but significant structure around Hokkaido corresponding to the Pacific Ocean and the Sea of Japan (Azuma *et al.* 2007). This structure is thought to be weak due to the high dispersal capability of the larvae. The ice goby, *Leucopsarion petersii*, shows not only two divergent lineages that diverged during the Pleistocene corresponding to the Pacific Ocean and the Sea of Japan but also morphological differences that also correspond to different selection patterns in these different regions.

Conclusions

The divergence between the eastern and western Pacific support the hypothesis of Pleistocene isolation into eastern and western populations. Although it is uncertain whether speciation has occurred in this taxon, the eastern and western Pacific populations have been evolving in isolation for at least 150,000 years. However, in addition to a clear history of an east-west population split, the western Pacific also contains relatively distinct and ancient lineages. This suggests that there may have been multiple glacial refugia in the western Pacific.

Numerous studies have stated many speciation events thought to have occurred as a result of the glacial cycles actually pre-dated the Pleistocene challenging and seeming to disprove the paradigm of Late Pleistocene speciation (Bush 1994; Cracraft & Prum 1988; Riddle 1996). However, more recent studies

such as this one are using multilocus data with model-based data allowing a more accurate estimate of divergence times than single locus sequence divergence, which tends to overestimate the age of speciation events. Given this caveat, the paradigm may not have been as sufficiently challenged or disproven as previously thought, however it is possible that Pleistocene rates of speciation may have only been higher in certain habitats, such as the boreal habitat or in areas directly affected by the glaciers (Weir & Schluter 2004). Given recent studies of marine habitats, there is evidence that divergence and speciation in the North Pacific Ocean for a wide variety of taxa as a result of the climate change in the Pleistocene is common (Canino *et al.* 2010; Foltz *et al.* 2008; Grant & Utter 1984; Harlin-Cognato *et al*. 2006; Kai *et al*. 2011; Shen *et al*. 2011).

Literature Cited

Addicott WO (1966) Late Pleistocene marine paleoecology and zoogeography in central California. *United States Geological Survey Professional Paper*, **523-C**, 1343-1346.

Arbogast BS, Edwards SV, Wakely J, Beerli P, Slowinski J (2002) Estimating divergence times from molecular data on phylogenetic and population genetic timescales. *Annual Review of Ecology and Systematics*, **33**, 707-740.

Azuma N, Kunihiro Y, Sasaki J, Mihara E, MIhara Y, Yasunaga T, Jin D, Abe S (2007) Genetic variation and population structure of hair crab (*Erimacrus isenbeckii*) in Japan inferred from mitochondrial DNA sequence analysis. *Marine Biotechnology*, **10**, 39-48.

Bennet KD, Tzedakis PC, Willis KJ (1991) Quaternary refugia of north European trees. *Journal of Biogeography*, **18**, 103-115.

Briggs JC (1974) Marine Zoogeography. McGraw-Hill, Toronto.

Buckley TR, Cordeiro M, Marshall DC, Simon C (2006) Differentiating between hypotheses of lineage sorting and introgression in New Zealand alpine cicadas (*Maoricicada* Dugdale). *Systematic Biology*, **55**, 411-425.

Bush MB (1994) Amazonian speciation: a necessarily complex model. *Journal of Biogeography*, **21**, 5 –17.

Canino MF, Spies IB, Cunningham KM, Hauser L, Grant WS (2010) Multiple ice-age refugia in Pacific cod, *Gadus macrocephalus. Molecular Ecology*, **19**, 4339-4351.

Carraro PE, Ager TA, Baichtal JF (2007) Possible refugia in the Alexander Archipelago of southeastern Alaska during the late Wisconsisn glaciation. *Canadian Journal of Earth Sciences*, **44**, 229-244.

Carstens BC, Knowles LL (2007) Shifting distributions and speciation: species divergence during rapid climate change. *Molecular Ecology*, **16**, 619-627.

Cassone BJ, Boulding EG (2006) Genetic structure and phylogeography of the lined shore crab, *Pachygrapsus crassipes*, along the northestern and western Pacific coasts. *Marine Biology*, **149**, 213-226.

Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD (2003) Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Research*, **31**, 3497-3500.

Clark PU, Ally RB, Pollard D (1999) Northern Hemisphere ice-sheet influences on global climate change. *Science*, **286**, 1104-1111.

Clement M, Posada D, Crandall K (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology* ,**9**, 1657-1659.

Coan EV (1971) The northwest American Tellinidae. *Veliger*, **14** (supplement): 1-63.

Collins TM, Frazer K, Palmer AR, Vermeij GJ, Brown WM (1996) Evolutionary history of northern hemisphere Nucella (Gastropoda, Muricidae): molecules, ecology, and fossils. *Evolution*, **50**, 2287–2304.

Coope GR (1995) Insect faunas in Ice Age environments: why so little extinction? Pages 55-74 ***in*** J. H. Lawton and R. M. May, editors. Extinction rates. Oxford University Press, Oxford, England.

Cracraft J, Prum RO (1988) Patterns and processes of diversification: speciation and historical congruence in some Neotropical birds. *Evolution*, **42**, 603 – 620.

Davis MB, Shaw RG (2001) Range shifts and adaptive responses to Quaternary climate change. *Science,* **292**, 673-679.

Drummond AJ, Rambaut A, Shapiro B & Pybus OG (2005) Bayesian Coalescent Inference of Past Population Dynamics from Molecular Sequences. *Molecular Biology and Evolution*, **22**, 1185-1192.

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology,* **7**, 214.

Dupanloup I, Schneider S, Excoffier L (2002) A simulated annealing approach to define the genetic structure of populations. *Molecular Ecology*, **11**, 2571–2581.

Ehrich D, Gaudeul M, Assefa A, Koch MA, Mummenhoff K, Nemomissa S, Consortium I, Brochmann C (2007) Genetic consequences of Pleistocene range shifts: contrast between the Arctic, the Alps and the East African mountains. *Molecular Ecology*, **16**, 2542-2559.

Excoffier L, Laval G, and Schneider S (2005) Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* **1**, 47-50.

Foltz DW, Nguyen AT, Kiger JR, Mah CL (2008) Pleistocene speciation of sister taxa in a North Pacific clade of brooding sea stars (*Leptasterias*). *Marine Biology,* **154**, 593-602.

Fu Y (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics,* **147**, 915-925.

Grant WS, Utter FM (1984) Biochemical population genetics of Pacific herring (Clupea pallasi). *Canadian journal of Fisheries and Aquatic Sciences,* **41**, 1083-1088.

Grosberg RK, Levitan DR, Cameron BB (1996) Characterization of genetic structure and genealogies using RAPD-PCR markers: a random primer for the novice and nervous. In: Molecular Zoology: Advances, Strategies, and Protocols (eds Ferraris JD,

Harlin-Cognato A, Bickham JW, Loughlin TR, Honeycutt RL (2006) Glacial refugia and the phylogeography of Stellar's sea lion (*Eumatopias jubatus*). *Journal of Evolutionary Biology,* **19**, 955-969.

Hedrick PW (2005) Genetics of Populations, 3rd ed. Jones and Bartlett Publishers, Sudbury, Massachusetts.

Hewitt GM (1999) Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society*, **68**, 87 –112.

Hewitt, GM (2004) Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions: Biological Sciences*, **359**, 183-195.

Hey J, Nielsen R (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis. Genetics,* **167**, 747-760.

Hudson RR, Kaplan NL (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics*, **111**, 147–164.

Huelsenbeck J, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetics trees. *Bioinformatics,* **17**, 754-755.

Ilves KL, Taylor EB (2008) Evolutionary and biogeographical patterns within the smelt genus *Hypomesus* in the North Pacific Ocean. *Journal of Biogeography,* **35**, 48-64.

Insua A, López-Piñón MJ, Freire R, Méndez J (2003) Sequence analysis of the ribosomal DNA internal transcribed spacer region in some scallop species (Mullosca: Bivalvia: Pectinidae). *Genome,* **46**, 595-604.

Jablonski D, Sepkowski JJ (1996) Paleobiology, community ecology, and scales of ecological pattern. *Ecology*, **77**, 1367-1378.

Jeffreys H (1961) Theory of Probability. Oxford University Press, Oxford.

Kai Y, Orr JW, Sakai K, Nakabo T (2011) Genetic and morphological evidence for cryptic diversity in the *Careproctus rastrinus* species complex (Liparidae) of the North Pacific. *Ichthyology Research*, **58**, 143-154.

Keppel G, Van Niel KP, Wardell-Johnson GW, Yates CJ, Byrne M, Mucina L, Schut A, Hopper SD, Franklin SE (2011) Refugia:identifying and understanding safe havens for biodiversity under climate change. *Global Ecology and Biogeography*, **20**: no. doi: 10.1111/j.1466-8238.2011.00686.x

Kitamura A, Omote H, Oda M (2000) Molluscan response to early Pleistocene rapid warming in the Sea of Japan. *Geology*, **28**, 723-726.

Klicka J, Zink RM (1997) The importance of recent ice ages in speciation: a failed paradigm. *Science*, **277**, 1666-1669.

Knowlton N (1993) Sibling species in the sea. *Annual Review of Ecology and Systematics,* **24**, 189-216.

Kokito T, Nohara K (2010) Phylogeography and historical demography of the anadromous fish *Leucopsarion petersii* in relation to geological history and oceanography around the Japanese Archipelago. *Molecular Ecology*, **20**, 143-164.

Lewontin RC (1974) The genetic basis of evolutionary change. Columbia University Press, New York.

Lindberg DR (1982) Taxonomic notes on members of the genus *Collisella* from the North Pacific Ocean, including a description of a new species from Alaska (Gastropoda: Acmaeidae). *Wasmann Journal of Biology,* **40**:48-58.

Lindberg DR, Marincovich JR (1988) New species of limpets from the Neogene of Alaska (Patellogastropoda: Mollusca). *Arctic*, **41**, 167-172.

Liu J, Gao T, Wu S, Zhang Y (2007) Pleistocene isolation in the Northwestern Pacific marginal seas and limited dispersal in a marine fish, *Chelon haematocheilus* (Temminck & Schlegel, 1845). *Molecular Ecology*, **16**, 275-288.

Lovette, IJ (2005) Glacial cycles and the tempo of avian speciation. *Trends in Ecology and Evolution*, **20**, 2, 57-59.

Mantel NA (1967) The detection of disease clustering and a generalized regression approach. *Cancer Res*earch, **27**, 209–220.

Marko PB (1998) Historical allopatry and the biogeography of speciation in the prosobranch snail genus *Nucella*. *Evolution*, **52**, 757-774.

Marko PB (2002) Fossil calibration of molecular clocks and the divergence times of geminate species pairs separated by the Isthmus of Panama. Molecular Biology and *Evolution*, **19**, 2005– 2021.

Marko PB (2004) 'What's larvae got to do with it?' Disparate patterns of post-glacial population structure in two benthic marine gastropods with identical dispersal potential. *Molecular Ecology*, **13**, 597-611.

Marko PB, Hoffman JM, Emme SA, McGovern TM, Keever CC, Cox LN (2010) The 'Expansion-Contraction' model of Pleistocene biogeography: rocky shores suffer a sea change? *Molecular Ecology*, **19**, 146-169.

Matsukuma A (1986) Cenozoic glycymeridid bivalves of Japan. *Palaeontological Society of Japan Special Paper*, **29**, 77-94.

McGovern TM, Keever CC, Saski CA, Hart MW, Marko PB (2010) Divergence genetics analysis reveals historical population genetic processes leading to contrasting phylogeographic patterns in co-distributed species. *Molecular Ecology*, **19**, 5043-5060.

Mitrovica JX (2003) Recent controversies in predicting post-glacial sea-level change. *Quaternary Science Reviews,* **22**, 127-133.

Nakano T, Sasaki T, Kase T (2010) Color polymorphism and historical biogeography in the Japanese patellogastropod limpet *Cellana nigrolineata* (Reeve) (Patellogastropoda: Nacellidae) *Zoological Science*, **27**, 811-820.

Newton MA, Raftery AE (1994) Approximate Bayesian inference with the weighted likelihood bootstrap (with discussion). *Journal of the Royal Statistical Society, Series B,* **56**, 3–48.

Palmer AR, Gayron SD, Woodruff DS (1990) Reproductive, morphological and genetic evidence for two cryptic species of Northeastern Pacific *Nucella*. *Veliger*, **33**, 325-338.

Posada D (2006). ModelTest Server: a web-based tool for the statistical selection of models of nucleotide substitution online. *Nucleic Acids Research,* **34**, W700-W703.

Posada D (2006) Collapse. http://darwin.uvigo.es/software/collapse.html

Posada D, Crandall KA (1998) Modeltest: testing the model of DNA substitution. *Bioinformatics,* **14**, 9, 817-818.

Rambaut A (2002) SE-AL v. 2.0a11: sequence alignment program. (http://tree.bio.ed.ac.uk/software/seal/)

Rambaut A, Drummond AJ (2009) Tracer v1.5, http://beast.bio.ed.ac.uk/Tracer

Ramos-Onsins SE, Rozas J (2002) Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution,* **19**, 2092-2100.

Rand AL (1948) Glaciation, an isolating factor in speciation. *Evolution*, **2**, 314-321.

Riddle BR (1996) The molecular phylogeographic bridge between deep and shallow history in continental biotas. *Trends in Ecology and Evolution*, **11**, 187–228.

Rohling EJ, Fenton M, Jorissen FJ, Bertrand P, Ganssen G, Caulet JP (1998) Magnitudes of sea-level lowstands of the past 500,000 years. *Nature*, **394**, 162-165.

Roy K, Valentine JW, Jablonski D, Kidwell SM (1996) Scales of climatic variability and time averaging in Pleistocene biotas: implications for ecology and evolution. *Trends in Ecology and Evolution*, **11**, 458-463.

Schmitt T (2007) Molecular biogeography of Europe: Pleistocene cycles and postglacial
trends. *Frontiers in Zoology*, **4**, 11.

Shen K, Jamandre BW, Hsu C, Tzeng W, Durand, J (2011) Plio-Pleistocene sea level and temperature fluctuations in the northwestern pacific promoted speciation in the globally-distributed flathead mullet *Mugil cephalus. BMC Evolutionary Biology*, **11**, 83.

Suchard MA, Weiss RE, Sinsheimer JS (2001) Bayesian selection of continuous-time Markov chain evolutionary models. *Molecular Biology and Evolution*, **18**, 1001–1013.

Swofford DL (2001) PAUP* (Phylogenetic Analysis Using Parsimony). Sinauer, Sunderland, Massachusetts.

Taberlet P, Fumagalli L, Wust-Saucy A-G, Cossons J-F (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, **7**, 453–464.

Tajima, F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics,* **125**, 585–593.

Tanaka S, Takahashi K (2005) Late Quaternary paleoceanographic changes in the Bering Sea and the western subarctic Pacific based on radiolarian assemblages. *Deep-Sea Research II*, **52**, 2131-2149.

Valentine, JW, Jablonski, D. (1993) In *Species Diversity in Ecological Communities* (eds. Ricklefs RE Schluter D), Univ. of Chicago Press, Chicago, 341-349.

Vermeij, GJ (1989) Geographical restriction as a guide to the causes of extinction: the case of the cold northern oceans during the Neogene. *Paleobiology*, **15**, 4, 335-356.

Vermeij GJ, Palmer AR, Lindberg DR (1990) Range limits and dispersal of molluscs in the Aleutian islands, Alaska. *The Veliger,* **33**, 346-354.

Vucetich JA, Waite TA (2003) Spatial patterns of demography and genetic processes across the species' range: null hypotheses for landscape conservation genetics. *Conservation genetics*, **4**, 639-645.

Weir JT, Schluter D (2004) Ice sheets promote speciation in boreal birds. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **271**, 1881-1887.

Woerner AE, Cox MP and Hammer MF (2007) Recombination-Filtered Genomic Datasets by Information Maximization. *Bioinformatics,* **23**, 1851-1853.

Zink RM, Slowinski JB (1995) Evidence from molecular systematics for decreased avian diversification in the Pleistocene epoch. *Proceedings of the National Academy of Sciences USA,* **92**, 5832–5835.